STIC-Biotech/ChemLib 54829

From: Fredman, Jeffrey
Sent: Wednesday, November 14, 2001 12:57 PM
To: STIC-Biotech/ChemLib
Cc: Bugaisky, Gabriele
Subject: FW: 09/215435

PLEASE RUSH!

I Approve.

Jeff Fredman

Edward Hart
Technical Info Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9203

-----Original Message-----
Fr m: Bugaisky, Gabriele
Sent: Wednesday, November 14, 2001 12:08 PM
To: Fredman, Jeffrey
Subject: 09/215435

This is an old est case that needs further search:

please search a) polynt encoding all of SEQ ID NO:177 and the following portions of SEQ ID NO:177: aa 1-24 and 25-99.

b)polynt encoding all of SEQ ID NO:179 and the following portions of SEQ ID NO:179: aa 1-23 and 24-121

c)polynt encodingall of SEQ ID NO:225 and the following portions of SEQ ID NO:225:aa 1-22 and 23-227.

originally only specific nt sequences were searched, not any nt encoding the protein.

please rush, if possible, as I have to reopen AF.

thanks, gabi b.

Gabriele E. Bugaisky
- **au 1653**
- **cm1-10d09; mailbox 9b01**
- **308-4201**

RECEIVED
NOV 14 2001
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 11/14/01
Date Completed: 11/15/01
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 9 -> reverse to NA
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: EV
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-177 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:38 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pr9:HSDJ601O1       +  341.00  638.54  2.5e-27   33458  ! AL109656 Human DNA sequence f
gb_htg9:AC023971       -  341.00  625.34  1.4e-26   153788  ! AC023971 Homo sapiens chromc
gb_in2:AF078161        +  141.00  249.48  1.2e-05   11008  ! AF078161 Manduca sexta lacuni
gb_pat2:E12656         +  135.50  264.37  1.7e-06   555  ! E12656 Synthetic DNA encoding u
gb_om:BTU35642         +  134.00  254.72  6.0e-06   1198  ! U35642 Bos taurus alpha1-micrc
gb_in3:CELC37C3        +  133.00  221.83  0.0004    42521  ! U64857 Caenorhabditis elegans
gb_htg1:AC006915       +  133.00  211.89  0.0014    134095  ! AC006915 Caenorhabditis eleg
gb_htg6:AC019993       -  132.50  220.00  0.0005    46847  ! AC019993 Drosophila melanogas
gb_in1:AC003053        +  132.50  215.29  0.0009    80689  ! AC003053 Drosophila melanogas
gb_in1:AC005149        -  132.50  213.52  0.0012    98942  ! AC005149 Drosophila melanogas
gb_in1:AE003579        -  132.50  203.80  0.0041    304383  ! AE003579 Drosophila melanoga
gb_pat2:I63559         +  131.00  247.06  1.6e-05   1455  ! I63559 Sequence 6 from patent
gb_om:MEELACP          +  130.00  255.84  5.2e-06   419  ! AJ000490 Macropus eugenii mRNA
gb_pr8:HSA1MICR        +  129.50  245.58  1.9e-05   1221  ! X04494 Human mRNA for alpha-1-
gb_pat2:I05217         +  129.50  245.51  1.9e-05   1232  ! I05217 Sequence 7 from Patent
gb_pr9:HSHCR           +  129.50  245.46  2.0e-05   1239  ! X04225 Human mRNA for protein
gb_pr9:HSAIMBG3        +  129.50  234.05  8.4e-05   4627  ! X54818 Human gene for alpha-1-
gb_pr10:HUMITILC08     +  129.50  232.68  0.0001    5423  ! M88249 Human inter-alpha-tryps
gb_htg19:AL137850      -  129.50  201.72  0.0053    193939  ! AL137850 Homo sapiens chromc
gb_htg3:AC011161       -  129.50  201.43  0.0055    200532  ! AC011161 Homo sapiens clone
gb_sy:SYNBOVTRYI       +  129.00  256.71  4.6e-06   301  ! M31335 Bovine synthetic colostr
gb_in2:CBRG45N02       +  129.00  213.82  0.0011    42726  ! AC084649 Caenorhabditis brigg
gb_pat1:A31024         +  127.00  256.66  4.6e-06   191  ! A31024 DNA fragment for bikunir
gb_pat1:A31025         -  127.00  256.66  4.6e-06   191  ! A31025 DNA fragment for bikunir
gb_pat1:A19212         +  127.00  249.88  1.1e-05   418  ! A19212 Human kunitz type protea
gb_pat2:I29002         +  127.00  249.88  1.1e-05   418  ! I29002 Sequence 17 from patent
gb_om:TVU34208         +  126.00  248.76  1.3e-05   378  ! U34208 Trichosurus vulpecula ea
gb_ro1:D50586          +  125.50  235.70  6.8e-05   1522  ! D50586 Mus musculus mRNA for P
gb_pat2:E13092         +  124.00  243.55  2.5e-05   435  ! E13092 DNA encoding a protease
gb_pr5:AK024817        +  124.00  239.10  4.4e-05   728  ! AK024817 Homo sapiens cDNA: FLJ
gb_pat2:I14875         +  124.00  236.53  6.1e-05   979  ! I14875 Sequence 1 from patent U
gb_pat2:I92685         +  124.00  236.53  6.1e-05   979  ! I92685 Sequence 1 from patent U
gb_pr10:HUMHTFP        +  124.00  236.53  6.1e-05   979  ! L27624 Homo sapiens tissue fact
gb_sts2:G31809         +  124.00  235.20  7.3e-05   1142  ! G31809 sWSS4055 Eric D. Green
gb_pr10:HUMPP5         +  124.00  235.20  7.3e-05   1142  ! D29992 Homo sapiens mRNA for p
gb_ro1:MMAMBP          +  124.00  234.92  7.6e-05   1180  ! X68680 M.musculus AMBP mRNA fc
gb_pr7:BC005330        +  124.00  234.75  7.7e-05   1203  ! BC005330 Homo sapiens, tissue
gb_ro1:MUSAMIATI       +  124.00  234.53  7.9e-05   1234  ! D28812 Mouse mRNA for alpha-1
gb_pr9:HSDJ447F3       +  124.00  196.72  0.0101    97385  ! AL050348 Human DNA sequence f
gb_om:OAU00163         +  123.50  250.49  1.0e-05   174  ! U00163 Ovis aries secretory prc
gb_om:AF241828         +  123.50  234.54  7.9e-05   1098  ! AF241828 Ovis aries kunitz dom
gb_pat1:A19210         +  123.00  241.91  3.1e-05   418  ! A19210 Human kunitz type pro
gb_pat2:I29001         +  123.00  241.91  3.1e-05   418  ! I29001 Sequence 15 from pate
gb_in3:CEW01F3         +  123.00  204.14  0.0039    32797  ! Z92815 Caenorhabditis eleg
gb_htg24:CEY44A6       +  123.00  184.26  0.0501    326074  ! Z98863 Caenorhabditis ele

seq_name: gb_pr9:HSDJ601O1

seq_documentation_block:
LOCUS       HSDJ601O1   33458 bp    DNA             PRI       04-APR-2001
DEFINITION  Human DNA sequence from clone RP4-601O1 on chromosome 20. Contains
            the gene for a novel protein with Kunitz/Bovine pancreatic trypsin
            inhibitor domain, the 5' end of a novel gene, part of the gene for
            a novel protein with WAP-type (Whey Acidic Protein) 'four-disulfide
            core' domain, ESTs and GSSs, complete sequence.
ACCESSION   AL109656
VERSION     AL109656.10  GI:10944958
KEYWORDS    HTG; Kunitz; WAP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 33458)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 21, 2000 this sequence version replaced gi:9795206.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-601O1 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-447F3 is at 33359 in this sequence.
            The true right end of clone RP4-688G8 is at 100 in this sequence.
            RP4-601O1 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
FEATURES              Location/Qualifiers
     source           1. .33458
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="20"
                      /clone="RP4-601O1"
                      /clone_lib="RPCI-4"
     repeat_region    3. .225
                      /note="L1M4 repeat: matches 3977. .4203 of consensus"
     repeat_region    225. .428
                      /note="L1M4 repeat: matches 4321. .4525 of consensus"
     repeat_region    429. .696
                      /note="AluJb repeat: matches 2. .302 of consensus"
     repeat_region    697. .778
```

```
                     /note="L1M4 repeat: matches 4525. .4606 of consensus"
     repeat_region   779. .1089
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   1090. .1146
                     /note="L1M4 repeat: matches 4606. .4662 of consensus"
     repeat_region   1165. .1293
                     /note="LTR42 repeat: matches 1. .134 of consensus"
     repeat_region   1351. .1730
                     /note="L1MB5 repeat: matches 5005. .5402 of consensus"
     repeat_region   1731. .2032
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   2056. .2547
                     /note="L1MB5 repeat: matches 5374. .5865 of consensus"
     repeat_region   2543. .2729
                     /note="L1MC/D repeat: matches 5428. .5609 of consensus"
     repeat_region   2730. .3022
                     /note="AluSg repeat: matches 1. .293 of consensus"
     repeat_region   3023. .3181
                     /note="L1MC/D repeat: matches 5609. .5769 of consensus"
     repeat_region   3361. .3895
                     /note="L1 repeat: matches 3599. .4167 of consensus"
     repeat_region   3986. .4281
                     /note="L2 repeat: matches 534. .834 of consensus"
     repeat_region   4526. .4828
                     /note="AluSx repeat: matches 1. .300 of consensus"
     misc_feature    5372. .5726
                     /note="match: GSS: Em:AQ076126"
     repeat_region   5394. .5470
                     /note="MIR repeat: matches 48. .142 of consensus"
     gene            7401. .7609
                     /gene="dJ601O1.3"
     CDS             <7401. .>7609
                     /gene="dJ601O1.3"
                     /note="possibly a pseudogene
                     match: proteins: Sw:Q14508 Sw:P00993 Sw:Q28631 Sw:P14730
                     Tr:O44131 Wp:C08G9"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="dJ601O1.3 (novel protein with WAP-type (Whey
                     Acidic Protein) 'four-disulfide core' domain)"
                     /protein_id="CAC36024.1"
                     /db_xref="GI:13559237"
                     /translation="FPGSSLGFLSSEHYSSLLVRPFSYLFCVPEYILEPPPCISAPEN
                     CTHLCTMQEDCEKGFQCCSSFCGIV"
     mRNA            complement(7623. .7780)
                     /gene="dJ688G8.4"
                     /note="Continues in Em:AL031671 as dJ688G8.4
                     match: ESTs: Em:AI808149 Em:AI808159"
                     /evidence=not_experimental
                     /product="dJ601O1.2 (a putative novel protein)"
     gene            complement(7623. .7780)
                     /gene="dJ688G8.4"
     repeat_region   8352. .8422
                     /note="L2 repeat: matches 2596. .2670 of consensus"
     repeat_region   8434. .8656
                     /note="MIR repeat: matches 9. .260 of consensus"
     repeat_region   9290. .9419
                     /note="MIR repeat: matches 47. .183 of consensus"
     misc_feature    complement(9433. .9899)
                     /note="match: GSS: Em:AQ476416"
     repeat_region   9538. .9830
                     /note="AluSg repeat: matches 1. .293 of consensus"
     misc_feature    9923. .10651
                     /note="match: GSS: Em:AQ478366"
     repeat_region   10453. .10647
                     /note="MER3 repeat: matches 1. .204 of consensus"
     misc_feature    complement(10590. .10982)
                     /note="match: GSS: Em:AQ223524"
     repeat_region   11007. .11080
                     /note="MER69 repeat: matches 2. .77 of consensus"
     repeat_region   11529. .11954
                     /note="MER69 repeat: matches 2070. .2511 of consensus"
     repeat_region   13092. .13158
                     /note="L2 repeat: matches 2666. .2732 of consensus"
     repeat_region   13202. .13331
                     /note="AluSg/x repeat: matches 183. .312 of consensus"
     repeat_region   13332. .15510
                     /note="L1MB1 repeat: matches 3916. .6171 of consensus"
     repeat_region   15534. .15832
                     /note="AluSg repeat: matches 1. .299 of consensus"
     repeat_region   15846. .16746
                     /note="L1M4 repeat: matches 3472. .4394 of consensus"
     repeat_region   16739. .18594
                     /note="L1PA11 repeat: matches 11. .1921 of consensus"
     repeat_region   18595. .18888
                     /note="AluSq repeat: matches 1. .295 of consensus"
     repeat_region   18889. .20972
                     /note="L1PA11 repeat: matches 1921. .3456 of consensus"
     repeat_region   20973. .21280
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   21281. .23911
                     /note="L1PA11 repeat: matches 3456. .6164 of consensus"
     repeat_region   23926. .24129
                     /note="L1M4 repeat: matches 4390. .4613 of consensus"
     repeat_region   24130. .24438
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   24439. .24513
                     /note="L1M4 repeat: matches 4613. .4683 of consensus"
     repeat_region   24790. .25008
                     /note="MIR repeat: matches 1. .242 of consensus"
     mRNA            join(25397. .25528,26926. .27103,28675. .28876)
                     /gene="dJ601O1.1"
                     /note="match: cDNAs: Em:AK020315
                     match: ESTs: Em:AW303354"
                     /evidence=not_experimental
                     /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                     pancreatic trypsin inhibitor domain)"
     gene            25397. .28876
                     /gene="dJ601O1.1"
     CDS             join(<25397. .25528,26926. .27103,28675. .28681)
                     /gene="dJ601O1.1"
                     /note="match: proteins: Sw:P08592 Tr:Q13793 Sw:P05067
                     Sw:P12023 Sw:Q95241 Tr:Q9TUI0 Tr:Q91963 Sw:Q02445
                     Tr:O54819 Tr:Q9Z2U8 Sw:P12111 Tr:Q9XWX5 Tr:O61893
                     Sw:P00974 Sw:P04815 Tr:O16701 Sw:P00975 Tr:O70160
                     Sw:P02760 Sw:P48307 Wp:C37C3A Sw:P00991 Sw:P00978
                     Sw:P00976 Sw:P13371 Sw:P04365 Tr:Q9VQT9 Tr:O95925
                     Sw:P00993 Sw:P10280 Tr:Q9TWF8 Tr:Q9TWG0 Tr:Q9TWF9
                     Sw:P26228 Sw:P00980 Sw:P00992 Sw:P00979 Sw:P81658
                     Sw:P25660 Sw:P19761 Sw:P24541 Tr:Q9U8G8 Tr:Q14594
                     Tr:O95428 Tr:O35536"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                     pancreatic trypsin inhibitor domain)"
                     /protein_id="CAC36025.1"
                     /db_xref="GI:13559238"
                     /translation="AACCTMKSAKLGFLLRFFIFCSLNTLLLGGVNKIAEKICGDLKD
                     PCKLDMNFGSCYEVHFRYFYNRTSKRCETFVFSGCNGNLNNFKLKIEREVACVAKYKP
                     PR"
     repeat_region   26066. .26228
                     /note="MIR repeat: matches 53. .205 of consensus"
     repeat_region   27253. .27579
                     /note="AluSp repeat: matches 1. .303 of consensus"
     repeat_region   27649. .27728

alignment_scores:
            Quality:  341.00              Length:     60
              Ratio:  5.683                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177 x HSDJ601O1   ..

 Align seg 1/1  to: HSDJ601O1  from: 1  to: 33458
```

```
    39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
       ||||||||||||||||||||||||||||||||||||||||||||||||||
 26925 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 26974

    55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
       ||||||||||||||||||||||||||||||||||||||||||||||||||
 26975 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 27024

    72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
       ||||||||||||||||||||||||||||||||||||||||||||||||||
 27025 CCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 27074

    89 ValAlaCysValAlaLysTyrLysProPro 98
       ||||||||||||||||||||||||||||||
 27075 GTAGCCTGTGTTGCAAAATACAAACCACCG 27104

seq_name: gb_htg9:AC023971

seq_documentation_block:
LOCUS       AC023971   153788 bp    DNA             HTG       07-APR-2000
DEFINITION  Homo sapiens chromosome 20 clone RP11-770N8 map 20, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC023971
VERSION     AC023971.2  GI:7523758
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 20, clone RP11-770N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 7, 2000 this sequence version replaced gi:7008875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6573
                Center clone name: 770_N_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145038 bases at least Q40
                Consensus quality: 148490 bases at least Q30
                Consensus quality: 150621 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 152588; sum-of-contigs
                Quality coverage: 5.2 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1333: contig of 1333 bp in length
            *     1334 1433: gap of      100 bp
            *     1434     3969: contig of 2536 bp in length
            *     3970 4069: gap of      100 bp
            *     4070     7211: contig of 3142 bp in length
            *     7212 7311: gap of      100 bp
            *     7312    13518: contig of 6207 bp in length
            *    13519 13618: gap of      100 bp
            *    13619    20876: contig of 7258 bp in length
            *    20877 20976: gap of      100 bp
            *    20977    28838: contig of 7862 bp in length
            *    28839 28938: gap of      100 bp
            *    28939    37590: contig of 8652 bp in length
            *    37591 37690: gap of      100 bp
            *    37691    45814: contig of 8124 bp in length
            *    45815 45914: gap of      100 bp
            *    45915    62601: contig of 16687 bp in length
            *    62602 62701: gap of      100 bp
            *    62702    81817: contig of 19116 bp in length
            *    81818 81917: gap of      100 bp
            *    81918   100273: contig of 18356 bp in length
            *   100274 100373: gap of      100 bp
            *   100374   125266: contig of 24893 bp in length
            *   125267 125366: gap of      100 bp
            *   125367   153788: contig of 28422 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20"
                     /clone="RP11-770N8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .3969
                     /note="assembly_fragment"
     misc_feature    4070. .7211
                     /note="assembly_fragment"
     misc_feature    7312. .13518
                     /note="assembly_fragment"
     misc_feature    13619. .20876
                     /note="assembly_fragment"
     misc_feature    20977. .28838
                     /note="assembly_fragment"
     misc_feature    28939. .37590
                     /note="assembly_fragment"
     misc_feature    37691. .45814
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    45915. .62601
                     /note="assembly_fragment"
     misc_feature    62702. .81817
                     /note="assembly_fragment
                     clone_end:T7
```

```
                    vector_side:right"
     misc_feature   81918. .100273
                    /note="assembly_fragment"
     misc_feature   100374. .125266
                    /note="assembly_fragment"
     misc_feature   125367. .153788
                    /note="assembly_fragment"
BASE COUNT    44583 a  30904 c  30568 g  46532 t   1201 others
ORIGIN


alignment_scores:
            Quality:  341.00              Length:     60
              Ratio:  5.683                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177 x AC023971/rev   ..

 Align seg 1/1  to reverse of: AC023971  from: 1  to: 153788

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  134123 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 134074

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  134073 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 134024

      72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
         ||||||||||||||||||||||||||||||||||||||||||||||||
  134023 CCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 133974

      89 ValAlaCysValAlaLysTyrLysProPro 98
         ||||||||||||||||||||||||||||||
  133973 GTAGCCTGTGTTGCAAAATACAAACCACCG 133944

seq_name: gb_in2:AF078161

seq_documentation_block:
LOCUS       AF078161    11008 bp    mRNA            INV       31-OCT-1999
DEFINITION  Manduca sexta lacunin mRNA, complete cds.
ACCESSION   AF078161
VERSION     AF078161.1  GI:6164594
KEYWORDS    .
SOURCE      tobacco hornworm.
  ORGANISM  Manduca sexta
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Sphingiodea; Sphingidae; Sphinginae; Manduca.
REFERENCE   1  (bases 1 to 11008)
  AUTHORS   Nardi,J.B., Martos,R., Walden,K.K., Lampe,D.J. and Robertson,H.M.
  TITLE     Expression of lacunin, a large multidomain extracellular matrix
            protein, accompanies morphogenesis of epithelial monolayers in
            Manduca sexta
  JOURNAL   Insect Biochem. Mol. Biol. 29 (10), 883-897 (1999)
  MEDLINE   99457716
   PUBMED   10528409
REFERENCE   2  (bases 1 to 11008)
  AUTHORS   Nardi,J., Martos,R., Walden,K., Lampe,D. and Robertson,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1998) Department of Entomology, University of
            Illinois, 505 S. Goodwin, Urbana, IL 61801, USA
FEATURES             Location/Qualifiers
     source          1. .11008
                     /organism="Manduca sexta"
                     /db_xref="taxon:7130"
     sig_peptide     195. .257
     CDS             195. .9791
                     /note="large multidomain extracellular matrix protein"
                     /codon_start=1
                     /product="lacunin"
                     /protein_id="AAF04457.1"
                     /db_xref="GI:6164595"
                     /translation="MVLCNLRSLLVAVIVLSNCIAWTASRHHYTHNVSGHRSRHRRQG
                     KGLYLSSSYVIPGGEGTGWGDWGDSTPCSRTCGGGVASQKRICLKFGPDGQPQCTGGD
                     TKYFSCETQDCPPGSTDFRAEQCSKYDDVEYRNIKYKWKPYTRGPNPCELNCMPHGER
                     FYYRHQLKVIDGTRCNDDSFDVCVNGTCQPVGCDMMLGSNAREDKCRRCRGNGKNCYT
                     TNGVLDTQDLIKGYNDMLLIPEGATTIVIMEVKPSNNYLALRAKNGTYYLNGDYHIDF
                     PRSMMIAGALWFYERSQQGFAAPDKLRCLGPTTEPLYLSLLLQSVNVGIEYEYSLPTE
                     RAPRPTQNYTWVHEHFTECSATCGGGFQTRNVTCRTQMDLELVDDTLCDEGLKPVTNQ
                     TCNNNVCPPQWVTGEWGSCSHRCGSDGTQSREIQCEKIIVNGFRSIVSDKECFDLLGP
                     KPEKFRKCNVNITCPTWFTGPWKPCDTLYGEGKQTRQVVCYQKNGRRIDVLDDSECTD
                     ERPESEQKCMIHSEERTDWVASEWSGCDNCFSTMRTRIAKCTTYDRKLVDDSFCAHHP
                     LPVLQEPCDAATLPPCDVQWYATQWSKCSSECGDGTQTRRVFCGIYVNNTVVEVEESK
                     CAGLTKYSEKRNCTVPKEKCPSLWYTAPWSKCSKECGGGEQSRRVLCLRGDQFVSDCS
                     GLNVPDSVQVCNTGPCNKDEIITLDSHSTPIMDEEDCDYEEYYPEIGLDFALAFSGRT
                     SPDLMLSDTPSADIDSGTETGSIDTDYITVSDDFGISDDGSGSGFSSFSTDFTSDTTD
                     DYSTDSTDSTVTESSLSPSSVSDLKHTPRSENGTVKPKAAIATKDTTTSETETSVTDL
                     DYSTIEDTDYTVTDSSTSSDETTGTDDKSSTTTDFSSSDSTITETYESGLTDSSTLKE
                     TSETATDTPGLSTTSGDETSASDKSKSTDISETTEKLDSTSYSATTESSTDESTVEFE
                     SSLYTEDTSASTSSELSSRDITDESSTPYTKDIEGTTASGSTESTFTDETTMSKVTEE
                     SSVAEEETTKTTITEEVSGTSESASINSDKTTMTTLSEDTGKTSVSEEITTEMTVTEE
                     TSETSPTEGTSDKTTMSTVSEETESSSVTEETTTETTVVENATDISSTEVTASDKTTM
                     TTMSEESEKTTEEATTEITVTKEVTESSSTETATSDKTISTLSEETGKTSVAEESSTTE
                     KVTETTVTTMPEETGKTITSEEITIKTTVTEEPTDVGSSEAITSDKTTVSTASEETGK
                     YSVSEEETVKTTVAEASTEPSSTEAITSDKTKMSTISEETGKTSVSEEITVKTTVTEE
                     STEPSSTEAITSEKTSVISTSEETGKTSVSEEVTVKTTVTDEATEITSTVSIETKETS
                     VSGSTEELSTQASSKIESPTTESGITSHTTESEESTVSTTEKGEVTSETTELTKSIVS
                     TETMLSTTEKEETVMSTTELISTEEFSSTTESTEETGSTTEISTEYTESTTASQIIEA
                     STTVAPETNETGTPASALPVEGLSTSTEGITTQESEVTVITTEKSKPGTTLRSPWSTP
                     THRRRCKPRKRTCSKTMFGCCPDRKTPAKGPYDEGCPRPKTCHESKFGCCHDGLSPAL
                     GPFEEGCPTISCTDTLFGCCLSDNKTAAEGNDQEGCPPPPPACKSTVFGCCADDETEA
                     RGPDKEGCPELTTSTELPSTTESEAQTEPESTSSSEATTTNILSSSIAYENCTTSEFG
                     CCFDQETPATGPQGEGCPCNSTEFGCCPDGVSPARGNDFEGCVITCNLSSYGCCPDGE
                     TPAHGPDDLGCCLLSAYGCCPDNRKPAVGPHLEGCGCQYSSFGCCPDNATVARGPNFQ
                     GCGCQYTEHGCCPDRHTEAAGPDYDGCGCHTYQFGCCPDGVTIAKGQNHQGCDCRDSQ
                     YGCCGDGKTPATGPEREGCDCATSEYGCCPDGLTEAKGHKFYGCSDLPENKQAACGLP
                     HDRGTCRNYSVYWYYDLEYGGCSRFWYGGCEGNGNKFATKEECEDVCVQPAPKDACNL
                     PKVKGACLGYNIRWYYDAEQEQCSQFIYGGCLGNANNYASLQLCQEQCQPERSEDQCR
                     LPIDRGSCSGNFGRWGFNSETRRCEQFMWGGCEGNTNRFSSEIACIQRCDPPGKPKLP
                     DLCTLPAAIGDCADYRERWYYDTREKSCQRFYYGGCAGNGNNFATQAECEGRCSEAKI
                     TTTVRPTEAHPLTEMCFMEKDPGPCTDTETRWVYDYKLGKCVTFEYGGCGGNRNNFPT
                     EEYCQYYCGTAQDICQLPMRSGPCTESLMRWFYDPSSDSCSQFTYGGCDGNDNRFETR
                     DDCESRCRTGSAQTTTTPTTSVTVAVDSRIAEQLPTECHISPSEEPCQRSGQVWYLDQ
                     RLGECVSHDNLDSGDSCRHTGVFPTQEACERTCGAFRNINVCRYDLDPGPCRTYEAKY
                     FFDKASRSCREFAYGGCHGGPNRFSTIDECQEVCRSEMDPCKQVVEPGDCTSRYVMWY
                     YDNVRDTCLQFIYGGCHGNENRFETLEDCERKCRQRPESVSTTTAFIATTTTTTTLAP
                     TLPPTEAEPIAEECRTPESLVPCGANITTYYYDQESRSCVQGDIGGCRYANSFRTEEE
                     CERRCGAFRDINVCEARLDPGPCRSSIPKFYWEPASGSCLQYSYGGCGGGANRFSTIE
                     ECEEICGNTGPDVVGPDCGPYEEECLQLVCPYGISRDPTPEGCTRCACVEAPVDCGPM
                     RVECEQMRCTYGMARTQDLNGCEKCSCLPDPCVRQRCPLGESCVAIAKRDPVTNDAQF
                     YAVCKPVNEVLPDCDRFNDECSRLRCEYGVQRTRTVDGCERCSCVQVEIDCEPLKLEC
                     ERLKCQYLVERQVGADGCERCRCSEHPCKNKECGPGDRCVPTAYLDPITQESRFSSEC
                     RPVNKTGYCPVEQASTTEYPCPNECVDDADCRGVGKCCARGCGRACAVPDAPTSAAPH
                     TTTHAPDLPQSPQVSQETEPEVAAEEGGKATLRCIFHGNPPPKITWRRGEITIDGDRD
                     RYRLLSDGSLEIVSVYRNDSGVYICVAENDLGTARQEIYLKVKVPVHTKLAAASSSLLT
                     AGSELVLECEVDGYPEPDVYWTKDGRRIIPNDNIRISGSSVSHLTVARVSLSDAGMYE
                     CHARNNYSSHYSYLQIAVQPLTIPAKCTDSPFFANCPLIVRSKFCQHKYYSKFCCKSC
                     LEAGQLSAHEVEMQADQPLVRK"
     misc_feature    381. .533
                     /note="Region: thrombospondin type I/properdin domain 1"
     misc_feature    534. .1187
                     /note="Region: PAL domain"
     misc_feature    1227. .1391
                     /note="Region: thrombospondin type I/properdin domain 2"
     misc_feature    1401. .1586
                     /note="Region: thrombospondin type I/properdin domain 3"
     misc_feature    1593. .1760
                     /note="Region: thrombospondin type I/properdin domain 4"
     misc_feature    1773. .1940
                     /note="Region: thrombospondin type I/properdin domain 5"
     misc_feature    1950. .2126
                     /note="Region: thrombospondin type I/properdin domain 6"
     misc_feature    2136. .2288
                     /note="Region: thrombospondin type I/properdin domain 7"
```

```
     misc_feature    2289. .4886
                     /note="Region: TES domain"
     misc_feature    4887. .4958
                     /note="Region: lagrin domain 1"
     misc_feature    4974. .5045
                     /note="Region: lagrin domain 2"
     misc_feature    5058. .5132
                     /note="Region: lagrin domain 3"
     misc_feature    5211. .5222
                     /note="Region: lagrin domain 4"
     misc_feature    5352. .5423
                     /note="Region: lagrin domain 5"
     misc_feature    5427. .5498
                     /note="Region: lagrin domain 6"
     misc_feature    5508. .5579
                     /note="Region: lagrin domain 7"
     misc_feature    5580. .5651
                     /note="Region: lagrin domain 8"
     misc_feature    5655. .5726
                     /note="Region: lagrin domain 9"
     misc_feature    5730. .5801
                     /note="Region: lagrin domain 10"
     misc_feature    5805. .5876
                     /note="Region: lagrin domain 11"
     misc_feature    5880. .5951
                     /note="Region: lagrin domain 12"
     misc_feature    5955. .6026
                     /note="Region: lagrin domain 13"
     misc_feature    6057. .6209
                     /note="Region: Kunitz proteinase inhibitor domain 1"
     misc_feature    6234. .6386
                     /note="Region: Kunitz proteinase inhibitor domain 2"
     misc_feature    6411. .6563
                     /note="Region: Kunitz proteinase inhibitor domain 3"
     misc_feature    6597. .6749
                     /note="Region: Kunitz proteinase inhibitor domain 4"
     misc_feature    6810. .6962
                     /note="Region: Kunitz proteinase inhibitor domain 5"
     misc_feature    6981. .7133
                     /note="Region: Kunitz proteinase inhibitor domain 6"
     misc_feature    7224. .7385
                     /note="Region: Kunitz proteinase inhibitor domain 7"
     misc_feature    7410. .7562
                     /note="Region: Kunitz proteinase inhibitor domain 8"
     misc_feature    7581. .7733
                     /note="Region: Kunitz proteinase inhibitor domain 9"
     misc_feature    7848. .7997
                     /note="Region: Kunitz proteinase inhibitor domain 10"
     misc_feature    8022. .8174
                     /note="Region: Kunitz proteinase inhibitor domain 11"
     misc_feature    8208. .8516
                     /note="Region: antistasin-like domain 1"
     misc_feature    8544. .8852
                     /note="Region: antistasin-like domain 2"
     misc_feature    8868. .8999
                     /note="Region: WAP/4 disulfide domain"
     misc_feature    9105. .9281
                     /note="Region: immunoglobulin domain 1"
     misc_feature    9378. .9557
                     /note="Region: immunoglobulin domain 2"
     misc_feature    9627. .9722
                     /note="Region: PLAC domain"
BASE COUNT     3109 a   2750 c   2738 g   2411 t
ORIGIN


alignment_scores:
            Quality:  141.00            Length:    118
              Ratio:  2.136               Gaps:      6
 Percent Similarity:  55.932   Percent Identity: 30.508

alignment_block:
 US-09-215-435-177 x AF078161   ..
```

```
 Align seg 1/1  to: AF078161  from: 1  to: 11008

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhe..............  12
         :::|||:::|||   ||||||:::   :::|||:::
    6249 GTCAAGGGAGCCTGTTTGGGCTACAATATTCGCTGGTACTACGACGCCGA 6298

      13 ...................PheIlePhe......CysSerLeuAsnThrL  21
                             ||||||:::      |||
    6299 GCAGGAGCAATGCAGTCAGTTCATCTATGGCGGTTGT.............. 6335

      21 euLeuLeuGlyGlyValAsnLysIleAlaGlu...LysIleCysGlyAsp  36
           ||||||:::::|||:::   |||:::   ::::::|||   :::
    6336 .....CTTGGCAATGCAAACAACTACGCCTCACTACAACTTTGCCAAGAG 6380

      37 ..................LeuLysAspProCysLysLeuAspMetAsnPh  47
                             :::|||   |||:::|||   ::::::
    6381 CAATGTCAACCTGAAAGGAGTGAAGATCAATGCCGTCTGCCAATTGACCG 6430

      47 eGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerL  64
          |||||||||               |||:::   :::|||      ::::
    6431 AGGGTCTTGTTCCGGCAACTTTGGCCGCTGGGGCTTCAACAGCGAAACGC 6480

      64 ysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsn  80
         ::|||||||||   |||::::::::|||||:::||||||   |||:::
    6481 GACGCTGCGAGCAGTTCATGTGGGGCGGTTGTGAGGGCAACACCAATAGG 6530

      81 PheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysPr  97
         |||            :::   |||:::||||||:::   :::      ||
    6531 TTC............AGCTCTGAGATCGCTTGCATCCAGCGTTGTGACCC 6568

      97 oPro 98
         ||||
    6569 ACCA 6572
```

```
seq_name: gb_pat2:E12656

seq_documentation_block:
LOCUS       E12656        555 bp    DNA             PAT       24-JUN-1998
DEFINITION  Synthetic DNA encoding urine trypsin inhibitor linked with SUC2
            signal.
ACCESSION   E12656
VERSION     E12656.1  GI:3251488
KEYWORDS    JP 1997047295-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Ooya,T., Morita,M., Tanabe,T., Miura,M. and Kobayashi,K. .
  TITLE     PRODUCTION OF PROTEIN
  JOURNAL   Patent: JP 1997047295-A 1 18-FEB-1997;
            GREEN CROSS CORP:THE
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1997047295-A/1
            PD   18-FEB-1997
            PF   08-AUG-1995 JP 1995224690
            PI   OOYA TOMOSUKE, MORITA MASANORI, TANABE TOSHIZUMI, MIURA
            MASAMI, PI   KOBAYASHI KAORU
            PC   C12P21/02,C12N1/19,C12N15/09,(C12P21/02,C12R1:84),(C12N1/19,
            PC   C12R1:84);
            CC   strandedness: Single;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .555
            FT                   /organism='Artificial sequences' FT
            sig_peptide     1. .60
            FT                   /product='SUC2 signal peptide' FT
            mat_peptide     61. .504
            FT                   /product='urine trypsin inhibitor'.
```

```
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      121 a    143 c    164 g    127 t
ORIGIN


alignment_scores:
            Quality:  135.50              Length:     99
              Ratio:  2.710                 Gaps:      3
 Percent Similarity:  50.505   Percent Identity: 36.364

alignment_block:
 US-09-215-435-177 x E12656   ..

 Align seg 1/1  to: E12656  from: 1  to: 555

       9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGl 25
         ||||||:::   |||:::|||                 ||||||   |||
       7 CTTTTGCAAGCTTTCCTTTTC.............CTTTTGGCTGGTTT 41

      25 yValAsnLysIleAlaGluLysIleCys..................... 34
         :::   ||||||:::      :::
      42 TGCAGCCAAAATATCTGCAGCTGTGCTACCCCAAGAAGAGGAAGGATCAG 91

      35 .....GlyAspLeu.....................LysAspProCysLys 42
              |||:::|||                    :::|||   |||:::
      92 GGGGTGGGCAACTGGTAACTGAAGTCACCAAGAAAGAAGATTCCTGCCAG 141

      43 LeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTy 59
         |||      :::   |||   |||      :::      ||||||||||
     142 CTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCTA 191

      59 rAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnG 76
         ||||   ||||||      ||||||||||   ::::::||||||   |
     192 TAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATGG 241

      76 lyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         |||||   |||||||||      :::   |||         ::::||
     242 GCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 288

 seq_name: gb_om:BTU35642

 seq_documentation_block:
 LOCUS       BTU35642     1198 bp    mRNA            MAM       05-SEP-1996
 DEFINITION  Bos taurus alpha1-microglobulin/bikunin mRNA, complete cds.
 ACCESSION   U35642
 VERSION     U35642.1  GI:1016297
 KEYWORDS    .
 SOURCE      cow.
   ORGANISM  Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
 REFERENCE   1  (bases 1 to 1198)
   AUTHORS   Lindqvist,A. and Akerstrom,B.
   TITLE     Bovine alpha 1-microglobulin/bikunin. Isolation and
             characterization of liver cDNA and urinary alpha 1-microglobulin
   JOURNAL   Biochim. Biophys. Acta, Gene Struct. Expr. 1306 (1), 98-106 (1996)
   MEDLINE   96201710
 REFERENCE   2  (bases 1 to 1198)
   AUTHORS   Lindqvist,A.
   TITLE     Direct Submission
   JOURNAL   Submitted (08-SEP-1995) Annika Lindqvist, Molecular Signalling,
             Lund University, P.O. Box 94, Lund, 221 00, Sweden
 FEATURES             Location/Qualifiers
      source          1. .1198
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /tissue_type="liver"
      CDS             104. .1162
                      /note="precursor protein for alpha1-microglobulin and
                      bikunin"
                      /codon_start=1
                      /product="alpha1-microglobulin/bikunin"
                      /protein_id="AAB07599.1"
                      /db_xref="GI:1016298"
                      /translation="MRSLSGLLLLLTACLAVNASSVPTLPDDIQVQENFDLSRIYGKW
                      FNVAVGSTCPWLKRFKEKMTMSTVVLIAGPTSKEISVTNTHRRKGVCESISGTYEKTS
                      ADGKFLYHKAKWNITMESYVVHTNYDEYAIFLTKKLSRRHGPTITVKLYGREPQLRES
                      LLEEFREVALGVGIPEDAIFTMPDRGECVPGEQDPVPTPLSRARRAVLTQEEEGSGAG
                      QPVTNFSKKADSCQLDYSQGPCLGLFKRYFYNGTSMACETFLYGGCMGNGNNFLSEKE
                      CLQTCRTVEACNLPIVQGPCRSYIQLWAFDAVKGKCVRFSYGGCKGNGNKFYSEKECK
                      EYCGIPGEADEELLRFSN"
 BASE COUNT      284 a    308 c    354 g    252 t
 ORIGIN


 alignment_scores:
             Quality:  134.00              Length:     43
               Ratio:  4.467                 Gaps:      0
  Percent Similarity:  69.767   Percent Identity: 55.814

 alignment_block:
  US-09-215-435-177 x BTU35642   ..

  Align seg 1/1  to: BTU35642  from: 1  to: 1198

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||:::||||||   :::   |||   |||      :::
     788 GATTCCTGCCAGCTGGACTACTCACAAGGGCCTTGCCTGGGGCTGTTCAA 837

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||   ||||||      ||||||||||||:::::::
     838 GAGGTATTTCTATAACGGTACATCCATGGCCTGTGAGACCTTCTTATATG 887

      72 erGlyCysAsnGlyAsnLeuAsnAsnPhe 81
         ::|||||   ||||||   |||||||||
     888 GCGGCTGCATGGGGAATGGCAACAACTTC 916

 seq_name: gb_in3:CELC37C3

 seq_documentation_block:
 LOCUS       CELC37C3    42521 bp    DNA             INV       08-JUL-1998
 DEFINITION  Caenorhabditis elegans cosmid C37C3.
 ACCESSION   U64857
 VERSION     U64857.1  GI:1465829
 KEYWORDS    .
 SOURCE      Caenorhabditis elegans.
   ORGANISM  Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE   1  (bases 1 to 42521)
   AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
             Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
             Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
             Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
             Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
             Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
             O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
             Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
             Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
             Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
             Wilkinson-Sproat,J. and Wohldman,P.
   TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
             elegans
   JOURNAL   Nature 368 (6466), 32-38 (1994)
   MEDLINE   94150718
 REFERENCE   2  (bases 1 to 42521)
   AUTHORS   Geisel,C. and Bradshaw,H.
   TITLE     The sequence of C. elegans cosmid C37C3
   JOURNAL   Unpublished (1998)
 REFERENCE   3  (bases 1 to 42521)
   AUTHORS   Waterston,R.
   TITLE     Direct Submission
```

```
  JOURNAL   Submitted (22-JUL-1996)
REFERENCE   4  (bases 1 to 42521)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO 63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone


                        NEIGHBORING COSMID INFORMATION:


            The 5' cosmid is ZK742, 200 bp overlap;3' cosmid is F40A3, 200 bp
            overlap.  Actual start of this cosmid is at base position 1 of
            CELC37C3; actual end is at 5024 of CELF40A3

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1. .42521
                        /organism="Caenorhabditis elegans"
                        /strain="Bristol N2"
                        /db_xref="taxon:6239"
                        /clone="C37C3"
                        /chromosome="V"
     gene               513. .15440
                        /gene="C37C3.6b"
     gene               513. .6558
                        /gene="C37C3.6a"
     CDS                join(513. .608,664. .878,1380. .1535,1583. .1734,1785. .2541,
                        2666. .2897,2976. .3097,3149. .4733,4801. .5584,5659. .5871,
                        6194. .6558)
                        /gene="C37C3.6a"
                        /note="similar to the DPTI/Kunitz family of inhibitors;
                        most similar to tissue factor pathway inhibitor precursor;
                        coded for by C. elegans cDNA yk84d8.5; coded for by C.
                        elegans cDNA yk117d1.5; coded for by C. elegans cDNA
                        yk88e5.3; coded for by C. elegans cDNA yk132d4.3; coded
                        for by C. elegans cDNA yk123e8.3; coded for by C. elegans
                        cDNA yk88e5.5; coded for by C. elegans cDNA yk22b2.5;
                        coded for by C. elegans cDNA yk132d4.5; coded for by C.
                        elegans cDNA yk123e8.5; coded for by C. elegans cDNA
                        CEESD32F; coded for by C. elegans cDNA CEESH31FC; coded
                        for by C. elegans cDNA CEESL67F; coded for by C. elegans
                        cDNA CEESL73F; coded for by C. elegans cDNA CEESP33F;
                        coded for by C. elegans cDNA CEESP94F; coded for by C.
                        elegans cDNA CESAC89F; coded for by C. elegans cDNA
                        yk10f9.3; coded for by C. elegans cDNA yk10f9.5; coded for
                        by C. elegans cDNA yk22b2.3; coded for by C. elegans cDNA
                        yk3e8.3; coded for by C. elegans cDNA yk3e8.5; coded for
                        by C. elegans cDNA yk41h8.5; coded for by C. elegans cDNA
                        yk5c8.5; coded for by C. elegans cDNA yk148d2.5; coded for
                        by C. elegans cDNA yk152f10.5"
                        /codon_start=1
                        /protein_id="AAC25867.1"
                        /db_xref="GI:3294501"
                        /translation="MRLLLFSAALLLCSVPTWAFSLSSFFGSDVAQKPYLHPNSPPER
                        DPASSRMKRQAYQVYVDGDVSVTVDKSGQKETGNWGPWVPENECSRSCGGGVQLEKRQ
                        CSGDCTGASVRYISCNLNACESGTDFRAEQCSKFNDEALDGNYHKWTPYKGKNKCELV
                        CKPESGNFYYKWADKVVDGTKCDSKSNDICVDGECLPVGCDGKLGSSLKFDKCGKCDG
                        DGSTCKTIEGRFDERNLSPGYHDIIKLPEGATNIKIQEARKSTNNLALKNGSDHFYLN
                        GNGLIQVEKEVEVGGTIFVYDDAEPETLSAQGPLSEELTVALLFRKGSRDTAIKYEFS
                        IPLEEEVDYMYKFDNWTPCSVSCGKGVQTRNLYCIDGKNKGRVEDDLCEENNATKPEF
                        EKSCETVDCEAEWFTGDWESCSSTCGDQGQQYRVVYCHQVFANGRRVTVEDGNCTVER
                        PPVKQTCNRFACPEWQAGPWSACSEKCGDAFQYRSVTCRSEKEGEEGKLLAADACPAD
                        EQEKFDTERTCNLGPCEGLTFVTGEWNLCTRCNDTEETREVTCKDSQGRAYPLEKCLV
                        DNSTEIPTDTRSCATQPPCEYEWTVSEWSKCTTECGHGHKTRRVICAIHQNGGLEVVD
                        EGHCQAEKPEGKTNCTNEEKCTGTWYTSSWSECTAECGGGSQDRVAVCLNYDKKPVPE
                        WCDEAVKPSEKQDCNVDDCPTCVDSEFGCCPDNSTFATGEFNFGCSNCSETEFGCCAD
                        NVTVATGPNSKGCEEFVESPLNLEADVANADAEASGDAPELCSVTNENGEAVDVECAT
                        IAPITALLGDGELIGNDTDASNETIHCSKTEFGCCPDWYTAASGKGNEGCPSFTLGGC
                        NETQFGCCHDDVTLARGANLEGCGEPSCAASLYGCCKDRKTIAFGPHYSGCERSSFPC
                        ELSDFGCCPDGETAALGKNGTGCGENCLTTKFGCCPDGKTTAKGSHNEGCGCEFAQYG
                        CCPDGKSVAKGAGFYGCPESCAQSQFGCCPDGKTRARGENKEGCPCQYTRYGCCPDGE
                        TTALGPRNDGCDNCRYAKHGCCPDGETKALGPDGAGCPPTTTPPFLMGGTVAPHKIAA
                        CNQTQESGTVCGAGYKLAWHYDTTEGRCNQFWYGGCGGNDNNFASQDMCETICVEPPG
                        KGRCYLPRVDGPLRCDQLQPRYYYDHSKKHCVAFWWRGCLGNANNFNSFEECSMFCKD
                        VGPYDAPTTAAPPPPPQQNAQQYLPTPEVQQIEIQSAEQPQPQQPQQQQQQQQQQPQQ
                        PRQSMEDICRSRQDAGPCETYSDQWFYNAFSQECETFTYGGCGGNLNRFRSKDECEQR
                        CFFVHGAQPSAARQEQAQPAAQPAQPAQPSNIVSPPQQSASPVVVPSNSKQRDACHLN
                        VDQGRCKGAFDSWYYEVATGSCVTFKYTGCGGNANRFASKDQCESLCVKPASEAASAG
                        IDGAAGINSVCDEAKDTGPCTNFVTKWYYNKADGTCNRFHYGGCQGTNNRFDNEQQCK
                        AACQNHKDACQLPKVQGPCSGKHSYYYYNTASHQCETFTYGGCLGNTNRFATIEECQA
                        RCPSKF"
CDS                 join(513. .608,664. .878,1380. .1535,1583. .1734,1785. .2541,
                    2666. .2897,2976. .3097,3149. .4733,4801. .5584,5659. .5871,
                    6194. .6547,7038. .7134,7881. .8009,8060. .8410,8493. .8571,
                    13650. .13741,14139. .14727,14940. .15440)
                    /gene="C37C3.6b"
                    /note="similar to the BPTI/Kunitz family of inhibitors;
                    most similar to tissue factor pathway inhibitor precursor
                    (TFPI); coded for by C. elegans cDNA yk257g8.5; coded for
                    by C. elegans cDNA yk257g8.3; coded for by C. elegans cDNA
                    yk185b12.3"
                    /codon_start=1
                    /protein_id="AAC25868.1"
                    /db_xref="GI:3294502"
                    /translation="MRLLLFSAALLLCSVPTWAFSLSSFFGSDVAQKPYLHPNSPPER
                    DPASSRMKRQAYQVYVDGDVSVTVDKSGQKETGNWGPWVPENECSRSCGGGVQLEKRQ
                    CSGDCTGASVRYISCNLNACESGTDFRAEQCSKFNDEALDGNYHKWTPYKGKNKCELV
                    CKPESGNFYYKWADKVVDGTKCDSKSNDICVDGECLPVGCDGKLGSSLKFDKCGKCDG
                    DGSTCKTIEGRFDERNLSPGYHDIIKLPEGATNIKIQEARKSTNNLALKNGSDHFYLN
                    GNGLIQVEKEVEVGGTIFVYDDAEPETLSAQGPLSEELTVALLFRKGSRDTAIKYEFS
                    IPLEEEVDYMYKFDNWTPCSVSCGKGVQTRNLYCIDGKNKGRVEDDLCEENNATKPEF
                    EKSCETVDCEAEWFTGDWESCSSTCGDQGQQYRVVYCHQVFANGRRVTVEDGNCTVER
                    PPVKQTCNRFACPEWQAGPWSACSEKCGDAFQYRSVTCRSEKEGEEGKLLAADACPAD
                    EQEKFDTERTCNLGPCEGLTFVTGEWNLCTRCNDTEETREVTCKDSQGRAYPLEKCLV
                    DNSTEIPTDTRSCATQPPCEYEWTVSEWSKCTTECGHGHKTRRVICAIHQNGGLEVVD
                    EGHCQAEKPEGKTNCTNEEKCTGTWYTSSWSECTAECGGGSQDRVAVCLNYDKKPVPE
                    WCDEAVKPSEKQDCNVDDCPTCVDSEFGCCPDNSTFATGEFNFGCSNCSETEFGCCAD
                    NVTVATGPNSKGCEEFVESPLNLEADVANADAEASGDAPELCSVTNENGEAVDVECAT
                    IAPITALLGDGELIGNDTDASNETIHCSKTEFGCCPDWYTAASGKGNEGCPSFTLGGC
                    NETQFGCCHDDVTLARGANLEGCGEPSCAASLYGCCKDRKTIAFGPHYSGCERSSFPC
                    ELSDFGCCPDGETAALGKNGTGCGENCLTTKFGCCPDGKTTAKGSHNEGCGCEFAQYG
                    CCPDGKSVAKGAGFYGCPESCAQSQFGCCPDGKTRARGENKEGCPCQYTRYGCCPDGE
                    TTALGPRNDGCDNCRYAKHGCCPDGETKALGPDGAGCPPTTTPPFLMGGTVAPHKIAA
                    CNQTQESGTVCGAGYKLAWHYDTTEGRCNQFWYGGCGGNDNNFASQDMCETICVEPPG
                    KGRCYLPRVDGPLRCDQLQPRYYYDHSKKHCVAFWWRGCLGNANNFNSFEECSMFCKD
                    VGPYDAPTTAAPPPPPQQNAQQYLPTPEVQQIEIQSAEQPQPQQPQQQQQQQQQQPQQ
                    PRQSMEDICRSRQDAGPCETYSDQWFYNAFSQECETFTYGGCGGNLNRFRSKDECEQR
                    CFFVHGAQPSAARQEQAQPAAQPAQPAQPSNIVSPPQQSASPVVVPSNSKQRDACHLN
                    VDQGRCKGAFDSWYYEVATGSCVTFKYTGCGGNANRFASKDQCESLCVKPASEAASAG
                    IDGAAGINSVCDEAKDTGPCTNFVTKWYYNKADGTCNRFHYGGCQGTNNRFDNEQQCK
                    AACQNHKDACQLPKVQGPCSGKHSYYYYNTASHQCETFTYGGCLGNTNRFATIEECQA
                    RCPKDDQTTTTSQPEELPSLPLVQEDPQPRPAFSLKQSFAHSRRRDAPFARSVSARHH
                    TPDSEEERVDCYAVPDPGSCGDYRLVWHYSATSNSCRQFYYGGCAGNTNRFETRDKCE
```

```
                TSCVAKIEERVESVSEASKSLEEVRLTDPRMDSHFGYHDPEVDQIEEEAEYVIVDTGA
                LPELCMLPEQRGSCYDNILRWRFDSEKSQCVTFMYSGCNPNANHFTSQETCERACGKW
                RNVAVCELPAEHGDCQLAIPRWYHDPKTSQCQMMMWTGCGGNGNAFSSKADCESLCRV
                ETLWSNNTDFCTLERSAGPCTDSISMWYFDSTHLDCKPFTYGGCRGNQNRFVSKEQCQ
                QSCRPGDTKSEDICTLRPEPGPCRLGLEKYFYDPVIQSCHMFHYGGCEGNANRFDSEL
                DCFRRCSSVKVEASESERVGQLTSASTPVIYIVNKTAIFVGNTFRIRCNSYGVLPITW
                YKNGGLLQFGSRITEENDDTLEIVDALTADAGVYTCIAGQDSTMSEGVEVVIKRLPGH
                RTTSRPMLTPSKNFSLGTPPTPSPSTVSTTPFRIYTPGSAPSDARVSRPTSNSCMDVG
                NASTCDLIVKNGLCGKKRYGTFCCHTCTRVHNFKF"
     gene       complement(8582. .12734)
                /gene="C37C3.7"
     CDS        complement(join(8582. .8744,8795. .8868,8914. .9194,
                9378. .9558,9607. .9722,10028. .10144,10473. .10730,
                10960. .11122,11912. .12280,12651. .12734))
                /gene="C37C3.7"
                /note="contains similarity to EGF-like domains
                (PS:PS00022)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC25862.1"
                /db_xref="GI:1465832"
                /translation="MTWSEKQEYCRIGSNDLNTCETCVGKGSNCFWCGGKTKRCMPFD

alignment_scores:
            Quality:  133.00            Length:     62
              Ratio:  3.500               Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177 x CELC37C3   ..

 Align seg 1/1  to: CELC37C3  from: 1  to: 42521

      34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
         |||   :::   ||||||   |||:::|||           |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
         |         |||      |||:::||||||   :::|||   :::|||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      67 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 83
         |||||||:::::::::|||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      84 LysIleGluArgGluValAlaCysValAlaLysTyr 95
            |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555

 seq_name: gb_htg1:AC006915

 seq_documentation_block:
 LOCUS       AC006915   134095 bp    DNA             HTG       24-FEB-1999
 DEFINITION  Caenorhabditis elegans clone Y97E10y, *** SEQUENCING IN PROGRESS
             ***, 1 unordered pieces.
 ACCESSION   AC006915
 VERSION     AC006915.1  GI:4263433
 KEYWORDS    HTG; HTGS_PHASE1.
 SOURCE      Caenorhabditis elegans.
   ORGANISM  Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE   1  (bases 1 to 134095)
   AUTHORS   Waterston,R.H.
   TITLE     The sequence of Caehorhabditis elegans clone
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 134095)
   AUTHORS   Waterston,R.H.
   TITLE     Direct Submission
   JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
 COMMENT     * NOTE: This is a 'working draft' sequence. It currently
             * consists of 1 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1   134095: contig of 134095 bp in length.
 FEATURES             Location/Qualifiers
      source          1. .134095
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /clone="Y97E10y"
 BASE COUNT    43230 a   24499 c   23702 g   42664 t
 ORIGIN


 alignment_scores:
             Quality:  133.00            Length:     62
               Ratio:  3.500               Gaps:      0
  Percent Similarity:  61.290   Percent Identity: 40.323

 alignment_block:
  US-09-215-435-177 x AC006915   ..

  Align seg 1/1  to: AC006915  from: 1  to: 134095

       34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
          |||   :::   ||||||   |||:::|||           |||   ||
    54670 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 54719

       50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
          |         |||      |||:::||||||   :::|||   :::|||||
    54720 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 54769

       67 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 83
          |||||||:::::::::|||||   ||||||   |||:::|||
    54770 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 54819

       84 LysIleGluArgGluValAlaCysValAlaLysTyr 95
             |||   ::::::   |||   :::|||:::
    54820 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 54855

  seq_name: gb_htg6:AC019993

  seq_documentation_block:
  LOCUS       AC019993    46847 bp    DNA             HTG       03-JAN-2000
  DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
              pieces.
  ACCESSION   AC019993
  VERSION     AC019993.1  GI:6664904
  KEYWORDS    HTG; HTGS_PHASE2.
  SOURCE      fruit fly.
    ORGANISM  Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  REFERENCE   1  (bases 1 to 46847)
    AUTHORS   Adams,M. and Venter,J.C.
    TITLE     Direct Submission
    JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
              Rockville, MD, USA
  COMMENT     This sequence was identified as CDM:10211667 by the submitter.
              For more information on this record e-mail to fly@celera.com.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
  FEATURES             Location/Qualifiers
       source          1. .46847
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
  BASE COUNT    13884 a   9574 c   9324 g   14065 t
```

```
ORIGIN


alignment_scores:
            Quality:  132.50              Length:    101
              Ratio:  2.366                 Gaps:      4
 Percent Similarity:  55.446   Percent Identity: 31.683

alignment_block:
 US-09-215-435-177 x AC019993/rev   ..

 Align seg 1/1  to reverse of: AC019993  from: 1  to: 46847

      12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGlyVal..... 26
         |||:::::::|||:::::::::      ||||||   :::|||:::
    5730 TTCATAGTATTCGCGGCTGTT...CTGCTCCTGGCAGCCGGAATTCGTGC 5684

      27 ..........AsnLysIleAlaGluLys..............IleCysG 35
                    :::   ::::::||||||              ::::::
    5683 AGCTCCCTCGGATGTCGTCGTGGAGAAGGCGGAGGAGCCCGCGGTGGCCA 5634

      35 lyAspLeuLysAsp..............................ProCysLysLeu 43
            |||   |||                                |||
    5633 AACCCCTGCCTGATGCCAGTGAGCTCACTGTTCCGGAGGACTGTCATCAG 5584

      44 AspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAs 60
            :::   |||   |||:::   :::   :::||||||  ||||
    5583 CCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTATCGCTACGCCTACAA 5534

      60 nArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyA 77
         |      ::::::   ||||||   ||||||::::::||||||   ||||
    5533 CGTGGATACACAATCCTGCGAGGAGTTCGTTTACGGTGGATGTGCCGGCA 5484

      77 snLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysValAla 93
         ||   |||||||||:::   |||   :::   |||   ||||||::::::
    5483 ACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAACAGGCATGTTTGGTA 5434

      94 Lys 94
         |||
    5433 AAG 5431

seq_name: gb_in1:AC003053

seq_documentation_block:
LOCUS       AC003053    80689 bp    DNA              INV       17-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS00216 (D160)), complete
            sequence.
ACCESSION   AC003053 AC002863 AC003283 AC003284 AC003285 AC003286 AC003287
            AC002864 AC003288 AC003289 AC003290 AC003291 AC003292 AC003293
            AC003294 AC003295 AC003296 AC003297
VERSION     AC003053.1  GI:2584830
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clone DS00216
            (D160)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 80689)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 24A3-24C2
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 80689)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library location: 24-3.
FEATURES             Location/Qualifiers
     source          1. .80689
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="24A3-24C2"
                     /clone="P1 DS00216 (D160)"
                     /note="This sequence has not changed since its original
                     submission on 11/04/1997.  It was resubmitted in order to
                     include all secondary accession numbers for the subclones
                     belonging to this clone."
BASE COUNT    23494 a  16620 c  16241 g  24334 t
ORIGIN


alignment_scores:
            Quality:  132.50              Length:    101
              Ratio:  2.366                 Gaps:      4
 Percent Similarity:  55.446   Percent Identity: 31.683

alignment_block:
 US-09-215-435-177 x AC003053   ..

 Align seg 1/1  to: AC003053  from: 1  to: 80689

      12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGlyVal..... 26
         |||:::::::|||:::::::::      ||||||   :::|||:::
    5964 TTCATAGTATTCGCGGCTGTT...CTGCTCCTGGCAGCCGGAATTCGTGC 6010

      27 ..........AsnLysIleAlaGluLys..............IleCysG 35
                    :::   ::::::||||||              ::::::
    6011 AGCTCCCTCGGATGTCGTCGTGGAGAAGGCGGAGGAGCCCGCGGTGGCCA 6060

      35 lyAspLeuLysAsp..............................ProCysLysLeu 43
            |||   |||                                |||
    6061 AACCCCTGCCTGATGCCAGTGAGCTCACTGTTCCGGAGGACTGTCATCAG 6110

      44 AspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAs 60
            :::   |||   |||:::   :::   :::||||||  ||||
    6111 CCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTATCGCTACGCCTACAA 6160

      60 nArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyA 77
         |      ::::::   ||||||   ||||||::::::||||||   ||||
    6161 CGTGGATACACAATCCTGCGAGGAGTTCGTTTACGGTGGATGTGCCGGCA 6210

      77 snLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysValAla 93
         ||   |||||||||:::   |||   :::   |||   ||||||::::::
    6211 ACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAACAGGCATGTTTGGTA 6260

      94 Lys 94
         |||
    6261 AAG 6263

seq_name: gb_in1:AC005149

seq_documentation_block:
```

```
LOCUS       AC005149    98942 bp    DNA             INV       20-JUN-1998
DEFINITION  Drosophila melanogaster DNA sequence (Pls DS03364 (D237) and
            DS02911 (D145)), complete sequence.
ACCESSION   AC005149 AC004339 AC004278 AC003204 AC003205 AC002788 AC002789
            AC002790 AC002791 AC002792 AC002793 AC002794 AC002795 AC002796
            AC003206 AC002797 AC002798 AC002799 AC003207 AC002800 AC002801
            AC002802 AC003208
VERSION     AC005149.1  GI:3242679
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clones DS03364
            (D237) and DS02911 (D145)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 98942)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 24A1-24C5
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 98942)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
            to drosophila@mhgc.lbl.gov.
            Library locations: 4_36, 31_31.
FEATURES             Location/Qualifiers
     source          1. .98942
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="24A1-24C5"
                     /clone="Pls DS03364 (D237) and DS02911 (D145)"
                     /note="DS03364 (D237) extends from bp 1 to bp 79652 and
                     DS02911 (D145) extends from bp 15716 to bp 98943"
BASE COUNT    29273 a  20900 c  20166 g  28603 t
ORIGIN


alignment_scores:
            Quality:  132.50            Length:    101
              Ratio:  2.366               Gaps:      4
 Percent Similarity:  55.446   Percent Identity: 31.683

alignment_block:
 US-09-215-435-177 x AC005149/rev   ..

 Align seg 1/1  to reverse of: AC005149  from: 1  to: 98942

     12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGlyVal..... 26
        |||::::::|||:::::::::      ||||||   :::|||:::
   2596 TTCATAGTATTCGCGGCTGTT...CTGCTCCTGGCAGCCGGAATTCGTGC 2550

     27 .........AsnLysIleAlaGluLys.............IleCysG 35
                 :::   ::::::||||||             ::::::
   2549 AGCTCCCTCGGATGTCGTCGTGGAGAAGGCGGAGGAGCCCGCGGTGGCCA 2500

     35 lyAspLeuLysAsp.....................ProCysLysLeu 43
           |||   |||                        |||
   2499 AACCCCTGCCTGATGCCAGTGAGCTCACTGTTCCGGAGGACTGTCATCAG 2450

     44 AspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAs 60
           :::   |||   |||:::   :::   :::|||||   |||||
   2449 CCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTATCGCTACGCCTACAA 2400

     60 nArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyA 77
        |      ::::::   ||||||   ||||||::::::||||||   ||||
   2399 CGTGGATACACAATCCTGCGAGGAGTTCGTTTACGGTGGATGTGCCGGCA 2350

     77 snLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysValAla 93
        ||   ||||||||||:::   |||   :::   |||   ||||||::::::
   2349 ACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAACAGGCATGTTTGGTA 2300

     94 Lys 94
        |||
   2299 AAG 2297

seq_name: gb_in1:AE003579

seq_documentation_block:
LOCUS       AE003579   304383 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386046 section 12
            of 16, complete sequence.
ACCESSION   AE003579 AE002638
VERSION     AE003579.2  GI:10727349
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 304383)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
```

```
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 304383)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7295765.
FEATURES             Location/Qualifiers
     source          1. .304383
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(14340. .14462,14669. .15409,15464. .>15868)
                     /gene="CG2772"
                     /product="CT9119"
                     /db_xref="FLYBASE:FBan0002772"
                     /db_xref="FLYBASE:FBgn0031533"
     gene            <14340. .>15868
                     /gene="CG2772"
                     /map="23F1-23F3"
                     /db_xref="FLYBASE:FBan0002772"
                     /db_xref="FLYBASE:FBgn0031533"
     CDS             join(14358. .14462,14669. .15409,15464. .15868)
                     /gene="CG2772"
                     /note="CG2772 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0002772"
                     /db_xref="FLYBASE:FBgn0031533"
                     /protein_id="AAF51110.1"
                     /db_xref="GI:7295808"
                     /translation="MLAKVLTILAIVASAGAADDFDPFIDIPFKRLKTSAERIAEHGY
                     PAESHFVETPDGYVLNVFRIPHSPKLNSNGNEGESEASRPVVLIMHGLFSCSDCFLLN
                     GPEDALPYNYADAGYDVWLGNARGNIYSRNNTRLNVKHPYFWKFSWHEIGSIDLPATI
                     DYILAETGQQSLHYVGHSQGCTSFFVMGSYRPEYNAKIKTAHMLAPPVYMGNSTEGLI
                     VSTAPLFGHHGIGSTLLENQVLLPQNAFIQRVLDTTCSNQPIMLSYCKTLAILWGGPE
                     IGNLNQTLLPQIAETHPAGVSSNQAIHYIQSYASNDFRLYDWGSKRNLEYYGVSEPPA
                     YDLTKITSELYLYYGLADGSANKQDISRLPDLLPNLALLHEVPDSTWGHLDFIFATEV
                     KRVINDLVLDYSKAYDARETNNEI"
     mRNA            join(18041. .18296,18504. .18607,18677. .19562,19620. .19772,
                     19836. .>19886)
                     /gene="CG2774"
                     /product="CT9125"
                     /db_xref="FLYBASE:FBan0002774"
                     /db_xref="FLYBASE:FBgn0031534"
     gene            <18041. .>19886
                     /gene="CG2774"
                     /map="23F1-23F3"
                     /db_xref="FLYBASE:FBan0002774"
                     /db_xref="FLYBASE:FBgn0031534"
     CDS             join(18114. .18296,18504. .18607,18677. .19562,19620. .19772,
                     19836. .19886)
                     /gene="CG2774"
                     /note="CG2774 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0002774"
                     /db_xref="FLYBASE:FBgn0031534"
                     /protein_id="AAF51109.1"
                     /db_xref="GI:7295807"
                     /translation="MEVESPEHTRDFAEVDINNGAATGSEDEEEEVPAPGSVTLDRNE
                     SDLFVSALSPSSIGDIHPLQEVLTDDGDYFISIVVSDPQKIGDGMGSYLAYKVTTKTN
                     IPKFKRSEFSTLRRFSDFLGIHDLLVGKYMRLGRIIPPAPSKNIIGSTKVKISPQQSE
                     PGTPMTQEWVEIRRAALERFVHRTAQHPVLRVDLDFMNFLESDQELPRSVNTSALSGA
                     GVIRLFNKVGETVNKITYKMDENDPWFDDKITEVESLDANLQKLHNAMKSLVTSRREL
                     SLLTGLVAKSAAMLSTCEEHTGLSRALSNLADVEEKIELLRSEQANSDFFILAEFIKD
                     YLGLFGAIKCIFHERVKAFQNWQYAQMQLSKRRENRGRFELANRADKLDQAQQEVDEW
                     QGKVQRCQQQFDDISAEIKREMERFELTRVKDFKVNIIKYIEDQMAHQQQIVSYWEAF
                     APFAREIV"
     mRNA            join(20807. .20990,21059. .21597,21748. .>21915)
                     /gene="CG12795"
                     /product="CT38238"
                     /db_xref="FLYBASE:FBan0012795"
                     /db_xref="FLYBASE:FBgn0031535"
     gene            <20807. .>21915
                     /gene="CG12795"
                     /map="23F1-23F3"
                     /db_xref="FLYBASE:FBan0012795"
                     /db_xref="FLYBASE:FBgn0031535"
     CDS             join(20936. .20990,21059. .21597,21748. .21915)
                     /gene="CG12795"
                     /note="CG12795 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012795"
                     /db_xref="FLYBASE:FBgn0031535"
                     /protein_id="AAF51108.1"
                     /db_xref="GI:7295806"
                     /translation="MEFPHLGQHCSEATCNRLDFLPVKCDSCDKVFCASHYSYDRHSC
                     PGAYKKNVQVPVCPLCREPVPTPPGVEPDVTVGQHIDQQCKSESKKIYTNRCNAKGCK
                     RKELIPVTCSQCRLNFCLRHRHTSDHDCKPVPASSTTSSRGGFQSIFKTSSDSRSMAA
                     QAAERRQNRRPASNSISNTNSRPRPVQATQVQNIQGNMSEDEALARALALSIMEQDDA
                     AESNRQAPAPSNAQQVAVGGGGNQQGNGKDKCLLS"
     mRNA            complement(join(<22242. .22619,22678. .23381,23475. .23604,
                     23663. .23823,23883. .24022,24076. .24563,24621. .25106,
                     25190. .>25354))
                     /gene="CG3248"
                     /product="CT10923"
                     /db_xref="FLYBASE:FBan0003248"
                     /db_xref="FLYBASE:FBgn0031536"
     gene            complement(<22242. .>25354)
                     /gene="CG3248"
                     /map="23F1-23F3"
                     /db_xref="FLYBASE:FBan0003248"
                     /db_xref="FLYBASE:FBgn0031536"
     CDS             complement(join(22242. .22619,22678. .23381,23475. .23604,
                     23663. .23823,23883. .24022,24076. .24563,24621. .25106,
                     25190. .25354))
                     /gene="CG3248"
                     /note="CG3248 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003248"
                     /db_xref="FLYBASE:FBgn0031536"
                     /protein_id="AAF51107.1"
                     /db_xref="GI:7295805"
                     /translation="MDDVERIQSENENLRKIRNRLMQWESKTDPLAALSIQQEEHLDV
                     LTNLWRDSGPSVGPTDSSAAATSQSGDDIRLPAEGLQNTNEFLLWFADVSAEIEQRGD
                     ADYHKYLQQLEQRKAECSHMLDQIAGAMERLGALCDEYDFVSQKTSALNTASEQLIEE
                     QERLQELSHEIQRRLHYFSQVELLNQRLQSPTLSVASEAFRECLNKIDECLNYIEENP
                     KFKDAAAYNVKYRQCLAKASGLVRNYVTSVINQATEATLHPKNNMPDASAALKAPDAA
                     FALYYGKYQTAAAKVKRVAQLIESRSEHSLDYAQLMADLQQHYLAQRASVMSPAVNLS
                     IQNVKVAHKGDHCSLTRSACAFLVHVCQDEQRLFYQFFSTGAPHLTVYLEGLCTILYD
                     TMRPFIIHINHLETLAEICSILRIEMLEEHVQQNPVALEAFATIAHQLLQDVQERLVF
                     RAHLYLQSDIQNFNPSSGDLAYPEKLEMMESIALSLQEPAPLRRSDSRNSMVSSVSSA
                     VETESVATAYTVKQMIRRTLVCLSRLYRCVDRPIFQGLSQEALKLCIQSVSHAAGKIS
                     ANKTPIDGELFEIKHLLILREQIAPFRVDFTVKETSLDFSKVKTAAFGLLQKRKQLFS
                     MGSNNALLEFLLEGTPQIKEHLLDSRKEVDRQLKSVCEKYIKDAVQMLVGPLITFLEK
                     AQSLLAQSTPATPQSPESTKASYVLRQSPWASPQQISSIIQETQRLIKAKLAVLQRSM
                     QLYLSNRDTEFIIFRPIRNNIIQSFVKLEQLLTTNGYSTDDMIITSCPSAEQVSILLS
                     SASILAAEGVASFAAAARKISTSSSVEGSTVGRKLSAMSNKSELVEEPKVEEVAGQIE
                     VVPAEVQPPAKIEEEEVTETQTQANSE"
     mRNA            join(25648. .25833,25904. .26518,26573. .28394,28452. .28603)
                     /gene="CG8843"
                     /product="CT9189"
                     /db_xref="FLYBASE:FBan0008843"
                     /db_xref="FLYBASE:FBgn0031537"
     gene            <25648. .>28603
                     /gene="CG8843"
                     /map="23F1-23F3"
                     /db_xref="FLYBASE:FBan0008843"
                     /db_xref="FLYBASE:FBgn0031537"
     CDS             join(25725. .25833,25904. .26518,26573. .28394,28452. .28590)
```

```
                     /gene="CG8843"
                     /note="CG8843 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008843"
                     /db_xref="FLYBASE:FBgn0031537"
                     /protein_id="AAF51106.1"
                     /db_xref="GI:7295804"
                     /translation="MAPQPVVTGLSPKEGPPGTRVIIRGEFLGTRVQDLIGLKICGSD
                     CLLSAEWKSPNKIIARTGPAKGKGDIIVTTLSGGVGTSTVQFRAYHETIGPLKESAVW

alignment_scores:
            Quality:  132.50              Length:    101
              Ratio:  2.366                 Gaps:      4
 Percent Similarity:  55.446   Percent Identity: 31.683

alignment_block:
 US-09-215-435-177 x AE003579/rev   ..

 Align seg 1/1  to reverse of: AE003579  from: 1  to: 304383

      12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGlyVal..... 26
         |||:::::::|||:::::::::     ||||||   :::|||:::
  265745 TTCATAGTATTCGCGGCTGTT...CTGCTCCTGGCAGCCGGAATTCGTGC 265699

      27 .........AsnLysIleAlaGluLys.............IleCysG 35
                    :::   ::::::||||||           ::::::
  265698 AGCTCCCTCGGATGTCGTCGTGGAGAAGGCGGAGGAGCCCGCGGTGGCCA 265649

      35 lyAspLeuLysAsp.....................ProCysLysLeu 43
            |||   |||                       |||
  265648 AACCCCTGCCTGATGCCAGTGAGCTCACTGTTCCGGAGGACTGTCATCAG 265599

      44 AspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAs 60
            :::   |||   |||:::   :::   :::||||||   ||||||
  265598 CCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTATCGCTACGCCTACAA 265549

      60 nArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyA 77
         |     ::::::   ||||||   ||||||::::::||||||   ||||
  265548 CGTGGATACACAATCCTGCGAGGAGTTCGTTTACGGTGGATGTGCCGGCA 265499

      77 snLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysValAla 93
         ||   ||||||||||:::   |||   :::   |||   ||||||:::::
  265498 ACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAACAGGCATGTTTGGTA 265449

      94 Lys 94
         |||
  265448 AAG 265446

 seq_name: gb_pat2:I63559

 seq_documentation_block:
 LOCUS       I63559       1455 bp    DNA             PAT       07-OCT-1997
 DEFINITION  Sequence 6 from patent US 5663143.
 ACCESSION   I63559
 VERSION     I63559.1  GI:2481132
 KEYWORDS    .
 SOURCE      Unknown.
   ORGANISM  Unknown.
             Unclassified.
 REFERENCE   1  (bases 1 to 1455)
   AUTHORS   Ley,A.Charles, Ladner,R.Charles, Guterman,S.Kosow,
             Roberts,B.Lindsay, Markland,W. and Kent,R.Baribault.
   TITLE     Engineered human-derived kunitz domains that inhibit human
             neutrophil elastase
   JOURNAL   Patent: US 5663143-A 6 02-SEP-1997;
 FEATURES             Location/Qualifiers
     source          1. .1455
                     /organism="unknown"
 BASE COUNT      337 a    316 c    366 g    436 t
 ORIGIN


 alignment_scores:
            Quality:  131.00              Length:     85
              Ratio:  2.911                 Gaps:      2
 Percent Similarity:  52.941   Percent Identity: 37.647

alignment_block:
 US-09-215-435-177 x I63559   ..

 Align seg 1/1  to: I63559  from: 1  to: 1455

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::      ::::::                ||       ::
      10 TTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTGGCGCCAAAGA 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
      60 AGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 109

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
            ||||||||||||||   ||||||      ||||||||||||   :::
     110 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 159

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGlu..... 86
         :::|||||   ||||||   |||||||||      :::   |||
     160 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 209

      87 .........ArgGluVal.....................AlaCysValA 93
                   |||   |||                    :::|||:::|
     210 GCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGAAAGTTGTTTAG 259

      93 laLys 94
         |||||
     260 CAAAA 264

 seq_name: gb_om:MEELACP

 seq_documentation_block:
 LOCUS       MEELACP       419 bp    mRNA            MAM       15-APR-1998
 DEFINITION  Macropus eugenii mRNA for early lactation protein (ELP).
 ACCESSION   AJ000490
 VERSION     AJ000490.1  GI:3059123
 KEYWORDS    early lactation protein; ELP gene.
 SOURCE      tammar wallaby.
   ORGANISM  Macropus eugenii
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
 REFERENCE   1  (bases 1 to 419)
   AUTHORS   Simpson,K.J.
   TITLE     Direct Submission
   JOURNAL   Submitted (17-JUL-1997) Simpson K.J., Division of Molecular Biology
             and Genetics, Victorian Institute of Animal Science, 475 Mickleham
             Rd, Attwood, Victoria, 3049, AUSTRALIA
   REMARK    Revised by [3]
 REFERENCE   2  (bases 1 to 419)
   AUTHORS   Simpson,K.J., Shaw,D. and Nicholas,K.R.
   TITLE     Asynchronous expression of a putative protease inhibitor gene in
             the mammary gland during lactation in the Tammar wallaby, Macropus
             eugenii
   JOURNAL   Unpublished
 REFERENCE   3  (bases 1 to 419)
   AUTHORS   Simpson,K.J.
   TITLE     Direct Submission
   JOURNAL   Submitted (14-APR-1998) Simpson K.J., Division of Molecular Biology
             and Genetics, Victorian Institute of Animal Science, 475 Mickleham
             Rd, Attwood, Victoria, 3049, AUSTRALIA
 COMMENT     On Apr 18, 1998 this sequence version replaced gi:3046757.
 FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Macropus eugenii"
                     /db_xref="taxon:9315"
     gene            1. .252
                     /gene="ELP"
     CDS             <1. .252
                     /gene="ELP"
```

```
                     /codon_start=1
                     /evidence=experimental
                     /product="early lactation protein"
                     /protein_id="CAA04128.1"
                     /db_xref="GI:3046758"
                     /db_xref="SPTREMBL:O62845"
                     /translation="EKCLDQIQVNSLENLSLLVPSLCLLPPVRGNCSSQILHYFYNTT
                     SRTCETFIYSGCNGNRNNFNSEEYCLKTCRRNKNRNNNN"
     source          1. .255
                     /organism="Macropus eugenii"
                     /db_xref="taxon:9315"
                     /tissue_type="lactating mammary gland"
     polyA_signal    370. .375
     polyA_site      386
BASE COUNT      139 a     98 c     72 g    110 t
ORIGIN


alignment_scores:
            Quality:  130.00             Length:     41
              Ratio:  4.333                Gaps:      0
 Percent Similarity:  73.171   Percent Identity: 56.098

alignment_block:
 US-09-215-435-177 x MEELACP   ..

 Align seg 1/1  to: MEELACP  from: 1  to: 419

      41 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 57
         |||   |||            |||:::|||   :::      :::::::||
      67 TGTCTGCTCCCACCTGTGAGGGGTAACTGCAGTTCTCAAATTCTTCACTA 116

      57 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 74
         ||||||||||   ||||||:::   ||||||||||||:::::|||||||
     117 TTTCTACAACACAACCTCTCGAACCTGTGAGACTTTCATCTACAGTGGCT 166

      74 ysAsnGlyAsnLeuAsnAsnPhe 81
         ||||||||||   |||||||||
     167 GCAATGGCAATAGGAACAATTTT 189

seq_name: gb_pr8:HSA1MICR

seq_documentation_block:
LOCUS       HSA1MICR     1221 bp    mRNA            PRI       30-MAR-1995
DEFINITION  Human mRNA for alpha-1-microglobulin and HI-30.
ACCESSION   X04494
VERSION     X04494.1  GI:24478
KEYWORDS    alpha-1-microglobulin; glycoprotein; proteinase inhibitor; serum
            protein; signal peptide.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1221)
  AUTHORS   Kaumeyer,J.F., Polazzi,J.O. and Kotick,M.P.
  TITLE     The mRNA for a proteinase inhibitor related to the HI-30 domain of
            inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin
            (protein HC)
  JOURNAL   Nucleic Acids Res. 14 (20), 7839-7850 (1986)
  MEDLINE   87040757
COMMENT     Data kindly reviewed (11-MAY-1987) by Polazzi J.O.
FEATURES             Location/Qualifiers
     source          1. .1221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     CDS             46. .1104
                     /note="precursor polypeptide"
                     /codon_start=1
                     /protein_id="CAA28182.1"
                     /db_xref="GI:24479"
                     /db_xref="SWISS-PROT:P02760"
                     /translation="MRSLGALLLLLSACLAVSAGPVPTPPDNIQVQENFNISRIYGKW
                     YNLAIGSTCPWLKKIMDRMTVSTLVLGEGATEAEISMTSTRWRKGVCEETSGAYEKTD
                     TDGKFLYHKSKWNITMESYVVHTNYDEYAIFLTKKFSRHHGPTITAKLYGRAPQLRET
                     LLQDFRVVAQGVGIPEDSIFTMADRGECVPGEQEPEPILIPRVRRAVLPQEEEGSGGG
                     QLVTEVTKKEDSCQLGYSAGPCMGMTSRYFYNGTSMACETFQYGGCMGNGNNFVTEKE
                     CLQTCRTVAACNLPIVRGPCRAFIQLWAFDAVKGKCVLFPYGGCQGNGNKFYSEKECR
                     EYCGVPGDGDEELLRFSN"
     sig_peptide     46. .102
                     /product="pot. signal peptide (AA -19 to -1)"
     mat_peptide     103. .1101
                     /product="alpha-1-microglobulin-HI-30 protein (AA 1-333)"
     mat_peptide     103. .651
                     /product="alpha-1-microglobulin"
     mat_peptide     661. .1095
                     /product="HI-30"
     misc_feature    724. .891
                     /note="domain I"
     misc_feature    892. .1095
                     /note="domain II"
     misc_feature    1188. .1192
                     /note="put. polyA signal"
BASE COUNT      296 a    322 c    363 g    240 t
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x HSA1MICR   ..

 Align seg 1/1  to: HSA1MICR  from: 1  to: 1221

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||              ::
     694 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 728

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     729 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 778

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||   :::
     779 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 828

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||   ||||||   |||||||||      :::   |||
     829 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 878

      88 uValAlaCys 91
            :::|||
     879 GCAGACCTGC 888

seq_name: gb_pat2:I05217

seq_documentation_block:
LOCUS       I05217       1232 bp                    PAT       02-DEC-1994
DEFINITION  Sequence 7 from Patent EP 0255011.
ACCESSION   I05217
VERSION     I05217.1  GI:591138
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1232)
  AUTHORS   Kaumeyer,J.F., Kotick,M.P. and Polazzi,J.O.
  TITLE     Human inter-alpha-trypsin inhibitor gene
  JOURNAL   Patent: EP 0255011-A2 7 03-FEB-1988;
FEATURES             Location/Qualifiers
     source          1. .1232
                     /organism="unknown"
BASE COUNT      300 a    323 c    365 g    244 t
```

ORIGIN

alignment_scores:
Quality: 129.50 Length: 70
Ratio: 3.083 Gaps: 2
Percent Similarity: 60.000 Percent Identity: 41.429

alignment_block:
US-09-215-435-177 x I05217 ..

Align seg 1/1 to: I05217 from: 1 to: 1232

```
     24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
        ||||||      |||:::::::::::::::|||               ::
    699 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 733

     38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
        :|||   |||:::|||      :::   |||   |||      :::
    734 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 783

     55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
          ||||||||||||||   ||||||      ||||||||||||   :::
    784 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 833

     72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
        :::||||||   ||||||   |||||||||      :::   |||
    834 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 883

     88 uValAlaCys 91
          :::|||
    884 GCAGACCTGC 893
```

OM of: US-09-215-435-177 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612	+	539.00	1246.72	1.8e-61	526	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738	+	529.00	1223.04	3.7e-60	539	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455	+	427.00	991.25	3.0e-47	267	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569	+	427.00	991.10	3.1e-47	271	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79088	+	140.00	312.99	1.8e-09	467	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79086	+	138.00	308.30	3.3e-09	467	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32013	+	133.00	250.18	5.8e-06	42521		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90070	+	133.00	250.18	5.8e-06	42521		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79087	+	131.50	291.77	2.8e-08	530	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35164	+	131.00	296.98	1.4e-08	285	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV40046	+	129.50	288.97	4.0e-08	441	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN81432	+	129.50	278.40	1.5e-07	1232		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT78950	+	129.50	278.21	1.6e-07	1255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77939	+	129.50	276.77	1.9e-07	1444		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ06878	+	127.00	291.72	2.8e-08	191	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46936	+	127.00	283.66	7.9e-08	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79081	+	126.00	282.52	9.1e-08	372	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT15870	+	125.00	285.49	6.2e-08	222	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79079	+	124.00	286.19	5.7e-08	165	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79080	+	124.00	277.83	1.7e-07	372	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79083	+	124.00	276.22	2.0e-07	435	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ90469	+	124.00	267.88	6.0e-07	979	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46935	+	123.00	274.29	2.6e-07	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79077	+	122.00	282.28	9.4e-08	153	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV86987	+	121.00	266.01	7.6e-07	593	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46934	+	120.00	267.26	6.5e-07	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ51001	+	120.00	258.82	1.9e-06	950	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46933	+	119.00	264.92	8.7e-07	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29024	+	118.50	268.95	5.2e-07	252	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29017	+	118.00	271.24	3.9e-07	180	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35169	+	118.00	267.90	6.0e-07	249	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV33175	+	118.00	264.52	9.2e-07	346	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32238	+	118.00	252.49	4.3e-06	1114		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT10945	+	117.00	252.27	4.4e-06	906	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ39947	+	117.00	249.15	6.6e-06	1228		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ10992	+	116.00	256.59	2.5e-06	474	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ10993	+	116.00	252.69	4.2e-06	693	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51316	+	116.00	249.86	6.0e-06	912	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51317	+	116.00	249.86	6.0e-06	912	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ81396	+	116.00	249.68	6.2e-06	928	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46644	+	116.00	249.50	6.3e-06	945		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46931	+	116.00	249.50	6.3e-06	945		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT16056	+	116.00	248.27	7.4e-06	106		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:AAN90108	+	116.00	245.23	1.1e-05	143		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32167	+	116.00	245.23	1.1e-05	143		

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612

seq_documentation_block:
ID AAX97612 standard; DNA; 526 BP.
XX
AC AAX97612;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 76.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; AAY35928.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 210; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 526 BP; 162 A; 100 C; 99 G; 165 T; 0 other;

alignment_scores:
Quality: 539.00 Length: 99
Ratio: 5.444 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-177 x AAX97612 ..

Align seg 1/1 to: AAX97612 from: 1 to: 526

```
      1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

     17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     72 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 121

     34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    122 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 171

     51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    172 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 221

     67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    222 AACTTTTGTCTTCTCCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 271

     84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
        |||||||||||||||||||||||||||||||||||||||||||||||
    272 AAATAGAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACCGAGG 318
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738

seq_documentation_block:
ID AAX97738 standard; DNA; 539 BP.
XX
AC AAX97738;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 303.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; AAY36054.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 376; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 539 BP; 165 A; 103 C; 97 G; 162 T; 12 other;

alignment_scores:
Quality: 529.00 Length: 99
Ratio: 5.343 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.980

alignment_block:
US-09-215-435-177 x AAX97738 ..

Align seq 1/1 to: AAX97738 from: 1 to: 539

```
      1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

     17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     82 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 131

     34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    132 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 181

     51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    182 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 231

     67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
        ||||||||||||||||:::|||||||||||||||||||||||||||||||
    232 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 281

     84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
        ||||||||||||||||:::|||||||||||||||||||||||||||||
    282 AAATAGAACGTGAAGTAKCCTGTGTTGCAAAATACAAACCACCGAGG 328
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455

seq_documentation_block:
ID AAC00455 standard; cDNA; 267 BP.
XX
AC AAC00455;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 453.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.

```
DR   P-PSDB; AAG00449.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 453; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 267 BP; 74 A; 57 C; 52 G; 84 T; 0 other;


alignment_scores:
           Quality:  427.00              Length:     79
             Ratio:  5.405                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.734

alignment_block:
 US-09-215-435-177 x AAC00455  ..

 Align seg 1/1  to: AAC00455  from: 1  to: 267

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 80

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 130

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 180

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 230

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 79
         |||||||||||||||:::|||||||||||||||||||
     231 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAAC 267

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569

seq_documentation_block:
ID   AAX51569 standard; cDNA; 271 BP.
XX
AC   AAX51569;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:148.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12791.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 253-254; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 271 BP; 75 A; 57 C; 55 G; 84 T; 0 other;


alignment_scores:
           Quality:  427.00              Length:     79
             Ratio:  5.405                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.734

alignment_block:
 US-09-215-435-177 x AAX51569   ..

 Align seg 1/1  to: AAX51569  from: 1  to: 271

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 131

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     132 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 181

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     182 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 231

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 79
         |||||||||||||||:::|||||||||||||||||||
     232 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAAC 268
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79088

seq_documentation_block:
ID   AAT79088 standard; DNA; 467 BP.
XX
AC   AAT79088;
XX
DT   12-NOV-1997  (first entry)
XX
DE   Insert from plasmid pHH341.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..441
FT                   /*tag=  a
FT                   /product= elastase inhibitor Ep7-d21-RPDF-52-55
FT   sig_peptide     7..66
FT                   /*tag=  b
FT   mat_peptide     67..438
FT                   /*tag=  c
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25938.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 38; 37pp; Japanese.
XX
CC   This is the nucleotide sequence of the insert for the plasmid pHH341
CC   which contains the sequence encoding the novel elastase specific
CC   inhibitor Ep7-d21-RPDF-52-55 (AAT79081). The inhibitor coding sequence
CC   is linked downstream of the yeast invertase (SUC2) signal peptide
CC   sequence. The modified protease inhibitors are targeted to the protease
CC   elastase, especially from neutrophils and can be used to treat diseases
CC   associated with elastase.  Modifications of the active site were done by
CC   site directed mutagenesis.
XX
SQ   Sequence 467 BP; 90 A; 121 C; 132 G; 124 T; 0 other;


alignment_scores:
            Quality:  140.00              Length:     83
              Ratio:  2.979                Gaps:      2
 Percent Similarity:  56.627   Percent Identity: 40.964

alignment_block:
 US-09-215-435-177 x AAT79088   ..

 Align seg 1/1  to: AAT79088  from: 1  to: 467

       9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGl 25
         |||||||:::   |||:::|||                |||||   |||
      13 CTTTTGCAAGCTTTCCTTTTC.............CTTTTGGCTGGTTT 47

      25 yValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysL 42
          :::   |||||||:::   :::              |||   |||:
      48 CGCTGCCAAGATATCTGCTAGA.............CCAGATTTCTGTC 82

      42 ysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPhe 58
         ::|||      :::   |||   |||      :::      |||||||||
      83 AACTGGGCTACTCGGCCGGTCCCTGCGTTGCTATGTTTCCTAGGTATTTC 132

      59 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs 75
         ||||||   ||||||      |||:::|||||||||:::::|||||
     133 TATAATGGTACATCCATGGCCTGTCAGACTTTCGTGTACGGCGGATGCAT 182

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
          ||||||.  |||||||||      :::   |||         :::|||
     183 GGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 231

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79086

seq_documentation_block:
ID   AAT79086 standard; DNA; 467 BP.
XX
AC   AAT79086;
XX
DT   12-NOV-1997  (first entry)
XX
DE   Insert from plasmid pHH337.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..441
FT                   /*tag=  a
FT                   /product= elastase inhibitor
FT   sig_peptide     7..66
FT                   /*tag=  b
FT   mat_peptide     67..438
FT                   /*tag=  c
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25936.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 30; 37pp; Japanese.
XX
CC   This is the nucleotide sequence of the insert for the plasmid pHH337
CC   which contains the sequence encoding the novel elastase specific
CC   inhibitor Ep1-d21-RPDF-52-55 (AAT79080). The inhibitor coding sequence
CC   is linked downstream of the yeast invertase (SUC2) signal peptide
CC   sequence. The modified protease inhibitors are targeted to the protease
CC   elastase, especially from neutrophils and can be used to treat diseases
CC   associated with elastase.  Modifications of the active site were done by
CC   site directed mutagenesis.
XX
SQ   Sequence 467 BP; 90 A; 123 C; 130 G; 124 T; 0 other;


alignment_scores:
            Quality:  138.00              Length:     83
              Ratio:  3.000                Gaps:      2
 Percent Similarity:  55.422   Percent Identity: 40.964
```

```
alignment_block:
 US-09-215-435-177 x AAT79086   ..

 Align seg 1/1  to: AAT79086  from: 1  to: 467

       9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGl 25
         |||||||::   |||:::|||              ||||||   |||
      13 CTTTTGCAAGCTTTCCTTTTC..............CTTTTGGCTGGTTT 47

      25 yValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysL 42
          :::   ||||||:::   :::                  |||   |||:
      48 CGCTGCCAAGATATCTGCTAGA..............CCAGATTTCTGTC 82

      42 ysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPhe 58
         ::|||      :::   |||   |||                |||||||||
      83 AACTGGGCTACTCGGCCGGTCCCTGCATCGCTTTCTTTCCTAGGTATTTC 132

      59 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs 75
         ||||||   ||||||      |||:::|||||||||:::::|||||
     133 TATAATGGTACATCCATGGCCTGTCAGACTTTCGTGTACGGCGGATGCAT 182

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
          ||||||   |||||||||      :::   |||         :::|||
     183 GGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 231

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32013

seq_documentation_block:
ID   AAZ32013 standard; DNA; 42521 BP.
XX
AC   AAZ32013;
XX
DT   10-JAN-2000  (first entry)
XX
DE   Human METH1 related EST U64857.
XX
KW   Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW   cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW   angiogenesis inhibitor; abnormal wound healing; inflammation;
KW   rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW   diabetic retinopathy; macula degeneration; haemangioma; detection;
KW   arterial-venous malformation; immune deficiency; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937660-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-US01313.
XX
PR   23-JAN-1998;   98US-0072298.
PR   28-AUG-1998;   98US-0098539.
XX
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM;
XX
DR   WPI; 1999-590684/50.
XX
PT   New isolated metalloprotease thrombospondin polypeptides, useful for
PT   treating hyperproliferative disorders, cancers or autoimmune disorders
PT   -
XX
PS   Disclosure; Page 256-283; 457pp; English.
XX
CC   AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC   metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC   respectively. METH1 and METH2 have been found to be potent inhibitors of
CC   angiogenesis both in vitro and in vivo. They can be used for treating
CC   cancer and other disorders related to angiogenesis including abnormal
CC   wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC   endometrial bleeding disorders, diabetic retinopathy, some forms of
CC   macula degeneration, haemangiomas, and arterial-venous malformations.
CC   They may be useful in treating deficiencies or disorders of the immune
CC   system, by activating or inhibiting the proliferation, differentiation,
CC   or mobilisation (chemotaxis) of immune cells. The etiology of these
CC   immune deficiencies or disorders may be genetic, somatic, such as
CC   cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC   toxins), or infectious. They can also be used to treat inflammatory
CC   conditions, both chronic and acute conditions. The products can also be
CC   used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC   AAY49511 represent sequences given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 42521 BP; 13493 A; 8055 C; 8365 G; 12608 T; 0 other;


alignment_scores:
            Quality:  133.00              Length:     62
              Ratio:  3.500                Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177 x AAZ32013   ..

 Align seg 1/1  to: AAZ32013  from: 1  to: 42521

      34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
         |||   :::   ||||||   |||:::|||            |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
         |         |||      |||:::||||||   :::|||   :::||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      67 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 83
         ||||||||:::::::::||||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      84 LysIleGluArgGluValAlaCysValAlaLysTyr 95
            |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90070

seq_documentation_block:
ID   AAC90070 standard; DNA; 42521 BP.
XX
AC   AAC90070;
XX
DT   19-MAR-2001  (first entry)
XX
DE   U64857 cDNA clone.
XX
KW   METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW   cancer therapy; benign tumour; ocular angiogenic disease;
KW   rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW   vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW   scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW   coronary collateral; cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW   plaque neovascularisation; telangiectasia; haemophiliac joint; EST;
KW   angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW   Crohn's disease; atherosclerosis; birth control; ss.
XX
OS   Unidentified.
XX
PN   WO200071577-A1.
XX
PD   30-NOV-2000.
XX
PF   25-MAY-2000; 2000WO-US14462.
XX
```

```
PR   25-MAY-1999;   99US-0318208.
PR   20-JUL-1999;   99US-0144882.
PR   10-AUG-1999;   99US-0147823.
PR   13-AUG-1999;   99US-0373658.
PR   22-DEC-1999;   99US-0171503.
PR   22-FEB-2000; 2000US-0183792.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
PA   (JONA/) JONAK Z L.
PA   (TRUL/) TRULLI S H.
PA   (FORN/) FORNWALD J A.
PA   (TERR/) TERRETT J A.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM,  Jonak ZL,  Trulli SH;
PI   Fornwald JA,  Terrett JA;
XX
DR   WPI; 2001-025136/03.
XX
PT   METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT   inhibit angiogensis in the treatment of disorders such as cancer,
PT   rheumatoid arthritis and psoriasis -
XX
PS   Claim 7; Pages 556-583; 768pp; English.
XX
CC   The present invention relates to human METH1 and METH2, (ME for
CC   metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC   The present sequence is an expressed sequence tag (EST) for METH. METH
CC   can be used for inhibiting angiogenesis in an individual, and for
CC   treating cancer, benign tumours, an ocular angiogenic disease,
CC   rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC   vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC   scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC   coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC   ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   fibromuscular dysplasia, wound granulation, Crohn's disease or
CC   atherosclerosis. METH can also be used in birth control. METH can also
CC   be used in diagnostic methods for the prognosis of cancer.
XX
SQ   Sequence 42521 BP; 13493 A; 8055 C; 8365 G; 12608 T; 0 other;


alignment_scores:
            Quality:  133.00              Length:     62
              Ratio:  3.500                 Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177 x AAC90070   ..

 Align seg 1/1  to: AAC90070  from: 1  to: 42521

      34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
         |||   :::   ||||||   |||:::|||              |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
         |         |||      |||:::||||||   :::|||   :::||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      67 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 83
         ||||||||:::::::::||||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      84 LysIleGluArgGluValAlaCysValAlaLysTyr 95
            |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79087

seq_documentation_block:
ID   AAT79087 standard; DNA; 530 BP.
XX
AC   AAT79087;
XX
DT   12-NOV-1997  (first entry)
XX
DE   Insert from plasmid pHH355.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..504
FT                   /*tag= a
FT                   /product= elastase inhibitor Epl-UTI-RPDF-52-55
FT   sig_peptide     7..66
FT                   /*tag= b
FT   mat_peptide     67..501
FT                   /*tag= c
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25937.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 34; 37pp; Japanese.
XX
CC   This is the nucleotide sequence of the insert for the plasmid pHH355
CC   which contains the sequence encoding the novel elastase specific
CC   inhibitor Epl-UTI-RPDF-52-55 (AAT79083). The inhibitor coding sequence
CC   is linked downstream of the yeast invertase (SUC2) signal peptide
CC   sequence. The modified protease inhibitors are targeted to the protease
CC   elastase, especially from neutrophils and can be used to treat diseases
CC   associated with elastase.  Modifications of the active site were done by
CC   site directed mutagenesis.
XX
SQ   Sequence 530 BP; 109 A; 138 C; 151 G; 132 T; 0 other;


alignment_scores:
            Quality:  131.50              Length:     99
              Ratio:  2.684                 Gaps:      3
 Percent Similarity:  49.495   Percent Identity: 36.364

alignment_block:
 US-09-215-435-177 x AAT79087   ..

 Align seg 1/1  to: AAT79087  from: 1  to: 530

       9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGl 25
         |||||||:::   |||:::|||              ||||||   |||
      13 CTTTTGCAAGCTTTCCTTTTC..............CTTTTGGCTGGTTT 47

      25 yValAsnLysIleAlaGluLysIleCys..................... 34
          :::   |||||||:::      :::
      48 CGCTGCCAAGATATCTGCTGCTGTGCTACCCCAAGAAGAGGAAGGATCAG 97
```

```
      35 .....GlyAspLeu.....................LysAspProCysLys 42
                 |||:::|||                          |||   |||:::
      98 GGGGTGGCCAACTGGTAACCGAAGTCACCAAGAGACCAGATTTCTGTCAA 147

      43 LeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTy 59
         |||      :::   |||   |||                |||||||||||
     148 CTGGGCTACTCGGCCGGTCCCTGCATCGCTTTCTTTCCTAGGTATTTCTA 197

      59 rAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnG 76
         ||||   ||||||      |||:::|||||||||:::::::||||||   |
     198 TAATGGTACATCCATGGCCTGTCAGACTTTCGTGTACGGCGGATGCATGG 247

      76 lyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         |||||   |||||||||      :::   |||         :::|||
     248 GCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 294

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35164

seq_documentation_block:
ID   AAT35164 standard; DNA; 285 BP.
XX
AC   AAT35164;
XX
DT   14-FEB-1997  (first entry)
XX
DE   ITI-D1 Kunitz domain fusion gene.
XX
KW   Aprotinin; Kunitz domain; human neutrophil elastase; hNE;
KW   connective tissue; alpha 1 protease inhibitor; API; neutrophil;
KW   alpha1 antitrypsin; respiratory disorder; cystic fibrosis;
KW   smokers emphysema; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..285
FT                   /*tag= a
FT                   /product= IIIsp::itiD1::matureIII (initial fragment)
FT                             fusion protein
FT   misc_feature    1..3
FT                   /*tag= b
FT                   /transl_except= GTG encodes Met
FT   sig_peptide     1..54
FT                   /*tag= c
FT                   /note= "M13 III protein signal peptide"
FT   mat_peptide     55..234
FT                   /*tag= d
FT                   /note= "ITI-D1 Kunitz domain"
FT   mat_peptide     235..285
FT                   /*tag= e
FT                   /note= "M13 III protein"
XX
PN   WO9620278-A2.
XX
PD   04-JUL-1996.
XX
PF   15-DEC-1995;   95WO-US16349.
XX
PR   16-DEC-1994;   94US-0358160.
XX
PA   (PROT-) PROTEIN ENG CORP.
XX
PI   Guterman SK,  Kent RB,  Ladner RC,  Ley AC,  Markland W;
PI   Roberts BL;
XX
DR   WPI; 1996-321851/32.
DR   P-PSDB; AAR99213.
XX
PT   New engineered inhibitors of human neutrophil elastase - contg.
PT   aprotinin-like Kunitz domain for treating, e.g. cystic fibrosis or
PT   other respiratory disorders
XX
PS   Example 1; Page 43; 105pp; English.
XX
CC   Genetically engineered human derived Kunitz domains can be used to
CC   inhibit human neutrophil elastase, an enzyme involved in the
CC   elimination of pathogens and the restructuring of connective tissue.
CC   In cases of reduction of the circulating alpha-1-protease inhibitor
CC   (API or alpha1 antitrypsin), or the inactivation of API by oxidation
CC   (smokers emphysema), extensive destruction of the lung tissue may
CC   result from uncontrolled elastolytic activity of human neutrophil
CC   elastase. Other respiratory disorders such as cystic fibrosis are
CC   thought to be caused by human neutrophil elastase release by
CC   neutrophils.  The genetically engineered human derived Kunitz
CC   domains can be used to treat such respiratory disorders. See
CC   AAR99146-R99211. Fusion genes were used in the production of the
CC   Kunitz domain derivatives.  This sequence is a fragment of a fusion
CC   gene that encodes the M13 III signal peptide, ITI-D1 Kunitz domain
CC   and the mature M13 III protein. Substitutions in the ITI-D1 domain
CC   can be made by use of BglI, EagI, NcoI, StyI, PstI and KasI
CC   restriction sites found within the sequence or by
CC   oligonucleotide-directed mutagenesis of single-stranded phage DNA.
CC   The resulting peptide derivatives can be expressed as intact human
CC   neutrophil elastase binding proteins or can be incorporated into
CC   larger proteins as a domain.
XX
SQ   Sequence 285 BP; 74 A; 67 C; 68 G; 76 T; 0 other;


alignment_scores:
            Quality:  131.00              Length:     85
              Ratio:  2.911                 Gaps:      2
 Percent Similarity:  52.941   Percent Identity: 37.647

alignment_block:
 US-09-215-435-177 x AAT35164   ..

 Align seg 1/1  to: AAT35164  from: 1  to: 285

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::      ::::::            |||      ::
      10 TTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTGGCGCCAAAGA 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
      60 AGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 109

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||||   :::
     110 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 159

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGlu..... 86
         :::||||||   ||||||   |||||||||      :::   |||
     160 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 209

      87 .........ArgGluVal.....................AlaCysValA 93
                    |||   |||                  :::|||:::|
     210 GCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGAAAGTTGTTTAG 259

      93 laLys 94
         |||||
     260 CAAAA 264

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV40046

seq_documentation_block:
ID   AAV40046 standard; cDNA to mRNA; 441 BP.
XX
AC   AAV40046;
XX
DT   09-NOV-1998  (first entry)
XX
DE   rUTI cDNA SEQ ID NO:28 from WO9829453 Example 10.
XX
KW   Drug; cell membrane-directed drug; phospholipid; lipid bilayer;
KW   cell cortex; blood coagulation; inflammation; immunological disorder;
```

```
KW   ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..441
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9829453-A1.
XX
PD   09-JUL-1998.
XX
PF   05-JAN-1998;   98WO-JP00002.
XX
PR   27-DEC-1996;   96JP-0359053.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Hasegawa T,  Kuriyama S;
XX
DR   WPI; 1998-388051/33.
DR   P-PSDB; AAW69522.
XX
PT   Drugs containing peptide(s) with specific affinity to
PT   phospholipid(s) - such as phosphatidyl serine, for treatment of
PT   blood coagulation, inflammatory and immunological disorders
XX
PS   Example 10; Page 84; 117pp; Japanese.
XX
CC   The present invention describes drug compositions which contain as
CC   an active component a peptide which has specific affinity to particular
CC   phospholipids (such as phosphatidyl ethanolamine or phosphatidyl
CC   serine), especially to phospholipids which constitute a lipid bilayer of
CC   cellular cortex and of which the concentration in the bilayer increases
CC   in cells which are abnormal (e.g. through injury, denaturation or
CC   activation). In particular, the peptide contains a sequence having
CC   phospholipid affinity and a structure of formula (I): (A1)a-(A2)b-(A3)c,
CC   where (A1) is one of two specific sequences (see AAW69516 and AAW69519),
CC   (A2) and (A3) are TRYLRIHPQSWVHQIALR, LRYLRIHPQSWVHQIALR (see AAW69517)
CC   or MEVLGCEAQNLY (see AAW69518); a = 0-5; b = 1-5, and c = 0-5. Preferred
CC   are the formulae A1-A2-A3, A2-A3, A2-A2-A3, A2-A2-A2-A3 or A2-A2
CC   (especially A2-A2-A3, A2-A2-A2-A3 or A2-A2). The sequence is linked to a
CC   peptide such as a blood factor, especially thrombo-modulin, urina-statin
CC   or membrane cofactor protein. The drugs are used for the treatment and
CC   prevention of diseases involving blood coagulation, inflammatory and
CC   immunological disorders. The present sequence encodes a protein from the
CC   present invention.
XX
SQ   Sequence 441 BP; 99 A; 119 C; 133 G; 90 T; 0 other;


alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                Gaps:       2
  Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AAV40046   ..

 Align seg 1/1  to: AAV40046  from: 1  to: 441

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||               ::
      34 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 68

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||     :::   |||  |||      :::
      69 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 118

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
          ||||||||||||||   ||||||      ||||||||||||   :::
     119 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 168

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||||   ||||||   |||||||||      :::   |||
     169 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 218

      88 uValAlaCys 91
          :::|||
     219 GCAGACCTGC 228

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN81432

seq_documentation_block:
ID   AAN81432 standard; cDNA; 1232 BP.
XX
AC   AAN81432;
XX
DT   06-DEC-1990  (first entry)
XX
DE   Sequence encoding new fusion protein contg. alpha-1-microglobulin (AMG)
DE   and the HI-30 region of inter-alpha-trypsin inhibitor (III) light chain.
XX
KW   Serine protease; enzyme; pancreatitis; atherosclerosis;
KW   chronic inflammation; therapy; elastase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..1109
FT                   /*tag= a
FT   mat_peptide     108..656
FT                   /*tag= b
FT                   /product=AMG
FT   mat_peptide     666..1100
FT                   /*tag= c
FT                   /product=HI-30
FT   misc_feature    921..926
FT                   /*tag= d
FT                   /note="Differs from the protein sequence of HI-30
FT                         purified from urine"
FT   misc_feature    1077..1079
FT                   /*tag= e
FT                   /note="Differs from the protein sequence of HI-30
FT                         purified from urine"
XX
PN   EP255011-A.
XX
PD   03-FEB-1988.
XX
PF   20-JUL-1987;   87EP-0110461.
XX
PR   29-JUL-1986;   86US-0891469.
XX
PA   (MILE ) MILES LABORATORIES INC.
XX
PI   Kaumeyer JF,  Kotick MP,  Polazzi JO;
XX
DR   WPI; 1988-030262/05.
DR   P-PSDB; AAP81110.
XX
PT   New DNA sequence coding for fusion protein contg. alpha-microglobulin -
PT   and inter-alpha-trypsin inhibitor, useful for treating excessive
PT   elastase prodn.
XX
PS   Disclosure; ; p; English.
XX
CC   A cDNA sequence encoding AMG in sequence with the HI-30 coding region of
CC   the light chain is claimed. The cDNA also contains a 5'-sequence coding
CC   for a secretory leader. 2 Arg residues separate the HI-30 and AMG regions
CC   and preceding AMG is a signal sequence. ITI is a serine protease,
CC   potentially used for treating excessive release of hydrolytic enzymes, eg
CC   elastase, in conditions such as pancreatitis, athersclerosis and chronic
CC   inflammation.
```

```
XX
SQ   Sequence 1232 BP; 299 A; 325 C; 364 G; 244 T; 0 other;


alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AAN81432   ..

 Align seg 1/1  to: AAN81432  from: 1  to: 1232

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||              ::
     699 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 733

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     734 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 783

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||||   :::
     784 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 833

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     834 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 883

      88 uValAlaCys 91
            :::|||
     884 GCAGACCTGC 893

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT78950

seq_documentation_block:
ID   AAT78950 standard; DNA; 1255 BP.
XX
AC   AAT78950;
XX
DT   11-NOV-1997  (first entry)
XX
DE   Anti-trypsin inhibitor UTI coding sequence.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             681..1124
FT                   /*tag= a
FT                   /product= protease inhibitor UTI
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25928.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 4-8; 37pp; Japanese.
XX
CC   This is the nucleotide sequence encoding the anti-trypsin inhibitor UTI.
CC   which is a protein containing 2 Kunitz domains.  The nucleotide sequence
CC   was used to construct a novel protease inhibitor by replacing the active
CC   site of the Kunitz domain 1 (see AAT79080, AAT79081 and AAT79083).
CC   The modified protease inhibitors are targeted to the protease elastase,
CC   especially from neutrophils and can be used to treat diseases associated
CC   with elastase.  Modifications of the active site were done by site
CC   directed mutagenesis.
XX
SQ   Sequence 1255 BP; 310 A; 329 C; 370 G; 246 T; 0 other;


alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AAT78950   ..

 Align seg 1/1  to: AAT78950  from: 1  to: 1255

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||              ::
     714 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 748

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     749 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 798

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||||   :::
     799 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 848

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     849 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 898

      88 uValAlaCys 91
            :::|||
     899 GCAGACCTGC 908

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77939

seq_documentation_block:
ID   AAC77939 standard; cDNA; 1444 BP.
XX
AC   AAC77939;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:333.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
```

```
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43730.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 885; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1444 BP; 362 A; 383 C; 422 G; 277 T; 0 other;


alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AAC77939   ..

 Align seg 1/1  to: AAC77939  from: 1  to: 1444

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||                ::
     883 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 917

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||       :::
     918 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 967

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
         ||||||||||||||||   ||||||      ||||||||||||   :::
     968 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 1017

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||||   ||||||   |||||||||      :::   |||
    1018 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 1067

      88 uValAlaCys 91
            :::|||
    1068 GCAGACCTGC 1077

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ06878

 seq_documentation_block:
 ID   AAQ06878 standard; DNA; 191 BP.
 XX
```

```
AC   AAQ06878;
XX
DT   05-MAR-1991  (first entry)
XX
DE   Encodes human bikunin domain 1-derived protease inhibitor.
XX
KW   bikunin; serine protease inhibitor; rheumatoid arthritis; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   mutation        57..59
FT                   /*tag=  a
FT                   /note= "substitute with codon for Leu, Ile, Val,
FT                    Arg, Phe, Tyr, Thr or Lys"
FT
FT   mutation        63..65
FT                   /*tag=  b
FT                   /note= "substitute with codon for Leu, Arg, Ile,
FT                    Val or Lys"
FT
XX
PN   EP401508-A.
XX
PD   12-DEC-1990.
XX
PF   01-MAY-1990;   90EP-0108284.
XX
PR   18-JAN-1990;   90DE-4001244.
PR   13-MAY-1989;   89DE-3915689.
XX
PA   (FARB ) BAYER AG.
XX
PI   Fritz H,  Gebhard W,  Das R;
XX
DR   WPI; 1990-369615/50.
DR   P-PSDB; AAR08291.
XX
PT   Protease-inhibitor variant derived from human bikunin - prodn.
PT   used as therapeutic for treatment of eg. rheumatoid arthritis
XX
PS   Disclosure; Fig 2; 29pp; German.
XX
CC   The 5' end overhangs the 3' end of the complementary strand by 5
CC   bases; the 5' end of the complementary strand overhangs the 3' end
CC   of the sense strand by AGCTG. The sequence encodes residues 19-77 of
CC   mature bikunin in which Met(36) and/or Met(38) are substituted by
CC   another naturally occurring amino acid. The sequence comprises domain
CC   1 of bikunin and is synthesised from two synthetic oligonucleotides.
CC   The substitutions are achieved by site-directed mutagenesis.
CC   Plasmids containing the sequences are used to transform bacteria
CC   and the protease inhibitory fragments are secreted as fusion
CC   proteins. The fragments can be used for the treatment of emphysema,
CC   septic shock, rheumatoid arthritis and coagulation disorders.
CC   See also AAQ06879.
XX
SQ   Sequence 191 BP; 49 A; 43 C; 46 G; 53 T; 0 other;


alignment_scores:
            Quality:  127.00              Length:     54
              Ratio:  3.848                Gaps:      0
 Percent Similarity:  61.111   Percent Identity: 46.296

alignment_block:
 US-09-215-435-177 x AAQ06878   ..

 Align seg 1/1  to: AAQ06878  from: 1  to: 191

      38 LysAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHi 54
         :::|||   |||:::|||      :::   |||   |||       :::
      18 GAAGACTCTTGCCAGCTGGGCTACTCTGCTGGTCCGTGCATGGGTATGAC 67

      54 sPheArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValP 71
            ||||||||||||||   ||||||      ||||||||||||   :
```

```
 68 TTCTCGTTACTTCTACAACGGTACCTCTATGGCTTGCGAAACTTTCCAGT 117

 71 heSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArg 87
    :::::||||||   ||||||   |||||||||      :::   |||
118 ACGGTGGTTGCATGGGTAACGGTAACAACTTCGTTACTGAAAAAGAATGC 167

 88 GluValAlaCys 91
        :::|||
168 CTGCAGACTTGC 179
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-177 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  3:58 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000


score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-6 +  131.00  293.43  1.1e-08   1455  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-17 +  127.00  296.56  7.2e-09   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-1 +  124.00  280.31  5.8e-08   979  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-1 +  124.00  280.31  5.8e-08   979  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-1 +  124.00  280.31  5.8e-08   979  !  S
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-817-145-2 +  124.00  278.73  7.1e-08   1140  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-15 +  123.00  286.71  2.6e-08   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-13 +  120.00  279.32  6.6e-08   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-11 +  119.00  276.86  9.0e-08   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-135 +  118.00  279.78  6.2e-08   249  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-2 +  116.00  268.16  2.8e-07   474  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-4 +  116.00  264.20  4.6e-07   693  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-026-145-1 +  116.00  261.17  6.7e-07   928  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-446-646-8 +  116.00  261.17  6.7e-07   928  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-7 +  116.00  260.98  6.9e-07   945  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-854-764-1 +  116.00  259.73  8.1e-07   1065  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09377-1 +  116.00  259.73  8.1e-07   1065
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-054-782-1 +  116.00  256.66  1.2e-06   1431  !
/cgn1_7/ptodata/1/ina/backfiles1.seq:5466783-1 +  116.00  256.66  1.2e-06   1431  ! Pate
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-6 +  115.00  263.00  5.3e-07   614  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-14 +  114.00  274.22  1.3e-07   165  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-14 +  114.00  274.22  1.3e-07   165  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-14 +  114.00  274.22  1.3e-07   165  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-70 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-74 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-70 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-74 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-70 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-74 +  113.00  262.23  5.9e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-58 +  111.00  257.15  1.1e-06   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-58 +  111.00  257.15  1.1e-06   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-58 +  111.00  257.15  1.1e-06   418  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-829-876-90 +  110.00  254.04  1.7e-06   445  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-685-660A-4 +  110.00  248.48  3.4e-06   759  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-974-196-4 +  110.00  248.48  3.4e-06   759  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-013-896A-1 +  110.00  240.65  9.4e-06   1610  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-985-692-1 +  107.50  249.02  3.2e-06   399  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-155-331-1 +  107.50  249.02  3.2e-06   399  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-424-022-1 +  107.50  249.02  3.2e-06   399  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-424-017B-1 +  107.50  249.02  3.2e-06   399  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-11696-1 +  107.50  249.02  3.2e-06   39
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-155-331-12 +  107.50  225.76  6.3e-05   3725
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-424-022-12 +  107.50  225.76  6.3e-05   3725
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-424-017B-12 +  107.50  225.76  6.3e-05   3725
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-11696-12 +  107.50  225.76  6.3e-05   3

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-6

seq_documentation_block:
; Sequence 6, Application US/08358160
; Patent No. 5663143
;  GENERAL INFORMATION:
;    APPLICANT:  LEY, Arthur C.
;    APPLICANT:  LADNER, Robert C.
;    APPLICANT:  GUTERMAN, Sonia K.
;    APPLICANT:  ROBERTS, Bruce L.
;    APPLICANT:  MARKLAND, William
;    APPLICANT:  KENT, Rachel B.
;    TITLE OF INVENTION:  ENGINEERED HUMAN-DERIVED KUNITZ
;    TITLE OF INVENTION:  DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  234
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W. Suite 300
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/358,160
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/133,031
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/009,319
;      FILING DATE:  26-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/664,989
;      FILING DATE:  01-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,063
;      FILING DATE:  02-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/240,160
;      FILING DATE:  02-SEP-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cooper, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  LEY=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1455 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  synthetic DNA fragment
US-08-358-160-6


alignment_scores:
            Quality:  131.00              Length:     85
```

```
            Ratio:  2.911                Gaps:      2
Percent Similarity:  52.941   Percent Identity: 37.647

alignment_block:
US-09-215-435-177 x US-08-358-160-6   ..

Align seg 1/1  to: US-08-358-160-6  from: 1  to: 1455

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::      ::::::             |||      ::
      10 TTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTGGCGCCAAAGA 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
      60 AGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 109

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||     ||||||||||||   :::
     110 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 159

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGlu..... 86
         :::|||||   ||||||   |||||||||      :::   |||
     160 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 209

      87 .........ArgGluVal.....................AlaCysValA 93
                 |||   |||                     :::|||:::|
     210 GCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGAAAGTTGTTTAG 259

      93 laLys 94
         |||||
     260 CAAAA 264

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-17

seq_documentation_block:
; Sequence 17, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-17


alignment_scores:
            Quality:  127.00              Length:     63
              Ratio:  3.175                 Gaps:      0
Percent Similarity:  63.492   Percent Identity: 41.270

alignment_block:
US-09-215-435-177 x US-08-321-658B-17   ..

Align seg 1/1  to: US-08-321-658B-17  from: 1  to: 418

      29 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 45
         :::||||||:::::                 |||   |||   |||:::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      45 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 62
          :::   |||   |||             :::|||||||||||||||:::
     262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

      62 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 78
         :::|||:::|||||   |||   ::::::|||||||:::|||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCAGGGGCAATATG 361

      79 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         ||||||||||||         |||   :::      |||
     362 AACAATTTTAAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-1

seq_documentation_block:
; Sequence 1, Application US/08147710
; Patent No. 5455338
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  4225 Roosevelt Way, N.E.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,710
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
```

```
;       REGISTRATION NUMBER:  31-684
;       REFERENCE/DOCKET NUMBER:  93-14
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-547-8080 ext 322
;       TELEFAX:  206-548-2329
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  979 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;       TISSUE TYPE:  Placenta
;     IMMEDIATE SOURCE:
;       CLONE:  J-2-11
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  39..746
US-08-147-710-1


alignment_scores:
            Quality:  124.00              Length:     86
              Ratio:  2.385                 Gaps:      1
 Percent Similarity:  60.465   Percent Identity: 31.395

alignment_block:
 US-09-215-435-177 x US-08-147-710-1   ..

 Align seg 1/1  to: US-08-147-710-1  from: 1  to: 979

      18 LeuAsnThrLeuLeuLeu.....GlyGlyValAsnLysIleAl 30
         |||:::   |||||||||           ::::::::::::      ||
      63 CTGTCGATTCTGCTGCTTTTCCTGACGGAGGCTGCACTGGGCGATGCTGC 112

      30 aGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnP 47
         |:::::::      |||:::      :::   |||   |||   :::::::
     113 TCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGACT 162

      47 heGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSer 63
         ::|||   |||      :::   :::|||||||:::|||:::|||   :::
     163 ACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACACG 212

      64 LysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAs 80
         :::   |||:::   |||:::::::::|||||||:::|||||||   |||||
     213 CAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAACAA 262

      80 nPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysP 97
         ||||                  :::   ||||||           |||
     263 TTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAAAG 312

      97 roProArg 99
            |||:::
     313 TTCCCAAA 320

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-1

seq_documentation_block:
; Sequence 1, Application US/08458090
; Patent No. 5728674
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,090
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  93-14D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Placenta
;    IMMEDIATE SOURCE:
;      CLONE:  J-2-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  39..746
US-08-458-090-1


alignment_scores:
            Quality:  124.00              Length:     86
              Ratio:  2.385                 Gaps:      1
 Percent Similarity:  60.465   Percent Identity: 31.395

alignment_block:
 US-09-215-435-177 x US-08-458-090-1   ..

 Align seg 1/1  to: US-08-458-090-1  from: 1  to: 979

      18 LeuAsnThrLeuLeuLeu.....GlyGlyValAsnLysIleAl 30
         |||:::   |||||||||           ::::::::::::      ||
      63 CTGTCGATTCTGCTGCTTTTCCTGACGGAGGCTGCACTGGGCGATGCTGC 112

      30 aGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnP 47
         |:::::::      |||:::      :::   |||   |||   :::::::
     113 TCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGACT 162

      47 heGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSer 63
         ::|||   |||      :::   :::|||||||:::|||:::|||   :::
     163 ACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACACG 212

      64 LysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAs 80
         :::   |||:::   |||:::::::::|||||||:::|||||||   |||||
     213 CAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAACAA 262

      80 nPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysP 97
         ||||                  :::   ||||||           |||
     263 TTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAAAG 312

      97 roProArg 99
            |||:::
     313 TTCCCAAA 320

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08457887
; Patent No. 5914315
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,887
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  93-14D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Placenta
;    IMMEDIATE SOURCE:
;      CLONE:  J-2-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  39..746
US-08-457-887-1


alignment_scores:
            Quality:  124.00              Length:     86
              Ratio:  2.385                 Gaps:      1
 Percent Similarity:  60.465   Percent Identity: 31.395

alignment_block:
 US-09-215-435-177 x US-08-457-887-1    ..

 Align seg 1/1  to: US-08-457-887-1  from: 1  to: 979

      18 LeuAsnThrLeuLeuLeu.....GlyGlyValAsnLysIleAl 30
         |||:::   |||||||||      :::::::::::      ||
      63 CTGTCGATTCTGCTGCTTTTCCTGACGGAGGCTGCACTGGGCGATGCTGC 112

      30 aGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnP 47
         |::::::     |||:::     :::   |||   |||   :::::::
     113 TCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGACT 162

      47 heGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSer 63
         ::|||   |||     :::   :::|||||:::|||:::|||   :::
     163 ACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACACG 212

      64 LysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAs 80
         :::   |||:::   |||::::::::::|||||:::||||||   |||||
     213 CAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAACAA 262

      80 nPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysP 97
         ||||                 :::   ||||||           |||
     263 TTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAAAG 312

      97 roProArg 99
         |||:::
     313 TTCCCAAA 320

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-817-145-2

seq_documentation_block:
; Sequence 2, Application US/08817145
; Patent No. 6025329
;  GENERAL INFORMATION:
;    APPLICANT:  UTSUMI, Jun
;    APPLICANT:  SUDO, Tetsuo
;    APPLICANT:  TANAKA, Yasuhiko
;    APPLICANT:  MATSUI, Mizuo
;    TITLE OF INVENTION:  THERAPEUTIC AGENT FOR OPHTHALMIC
;    TITLE OF INVENTION:  DISEASES
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch, LLP.
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/817,145
;      FILING DATE:  02-JUL-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  760-230P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..753
US-08-817-145-2


alignment_scores:
            Quality:  124.00              Length:     86
              Ratio:  2.385                 Gaps:      1
 Percent Similarity:  60.465   Percent Identity: 31.395

alignment_block:
 US-09-215-435-177 x US-08-817-145-2    ..

 Align seg 1/1  to: US-08-817-145-2  from: 1  to: 1140

      18 LeuAsnThrLeuLeuLeu.....GlyGlyValAsnLysIleAl 30
         |||:::   |||||||||      :::::::::::      ||
```

```
     73 CTGTCGATTCTGCTGCTTTTCCTGACGGAGGCTGCACTGGGCGATGCTGC 122

     30 aGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnP 47
        |:::::::      |||:::      :::   |||   |||   :::::::
    123 TCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGACT 172

     47 heGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSer 63
        ::|||   |||      :::   :::|||||||:::|||:::|||   :::
    173 ACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACACG 222

     64 LysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAs 80
        :::   |||:::   |||::::::::::|||||||:::|||||   |||||
    223 CAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAACAA 272

     80 nPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysP 97
        ||||            :::   ||||||           |||
    273 TTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAAAG 322

     97 roProArg 99
          |||:::
    323 TTCCCAAA 330

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-15

seq_documentation_block:
; Sequence 15, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-15


alignment_scores:
            Quality:  123.00              Length:     63
              Ratio:  3.075                Gaps:      0
 Percent Similarity:  63.492   Percent Identity: 39.683

alignment_block:
 US-09-215-435-177 x US-08-321-658B-15   ..

 Align seg 1/1  to: US-08-321-658B-15  from: 1  to: 418

     29 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 45
        :::||||||:::::              |||   |||   |||:::
    212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

     45 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 62
        :::   |||   |||            :::|||||||||||||||:::
    262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

     62 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 78
        :::|||:::||||||   |||   ::::::||||||:::||||||:::
    312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCAGGGGCAATATG 361

     79 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
        |||||||||:::         |||   :::       |||
    362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-13

seq_documentation_block:
; Sequence 13, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-13


alignment_scores:
            Quality:  120.00            Length:     63
              Ratio:  3.077               Gaps:      0
 Percent Similarity:  61.905   Percent Identity: 39.683

alignment_block:
 US-09-215-435-177 x US-08-321-658B-13   ..

 Align seg 1/1  to: US-08-321-658B-13  from: 1  to: 418

      29 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 45
         :::||||||::::::                 |||   |||   |||:::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      45 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 62
          :::   |||   |||              :::||||||||||||||:::
     262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

      62 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 78
         :::|||:::|||||||   |||   ::::::||||||   |||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCCTGGGCAATATG 361

      79 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         ||||||||||:::         |||   :::       |||
     362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

 seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-11

 seq_documentation_block:
 ; Sequence 11, Application US/08321658B
 ; Patent No. 5576294
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  No. 5576294ris, Fanny
 ;    APPLICANT:  No. 5576294ris, Kjeld
 ;    APPLICANT:  Bjorn, Soren E
 ;    APPLICANT:  Petersen, Lars C
 ;    APPLICANT:  Olsen, Ole H
 ;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
 ;    TITLE OF INVENTION:  VARIANT
 ;    NUMBER OF SEQUENCES:  18
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
 ;      STREET:  405 Lexington Avenue, 64th Floor
 ;      CITY:  New York
 ;      STATE:  New York
 ;      COUNTRY:  USA
 ;      ZIP:  10174-6401
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/321,658B
 ;      FILING DATE:  12-OCT-1994
 ;      CLASSIFICATION:  435
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Lowney Dr., Karen A.
 ;      REGISTRATION NUMBER:  31,274
 ;      REFERENCE/DOCKET NUMBER:  3693.210-US
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  212-867-0123
 ;      TELEFAX:  212-878-9655
 ;  INFORMATION FOR SEQ ID NO:  11:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  418 base pairs
 ;      TYPE:  nucleic acid
 ;      STRANDEDNESS:  single
 ;      TOPOLOGY:  linear
 ;    MOLECULE TYPE:  cDNA
 ;    ORIGINAL SOURCE:
 ;      ORGANISM:  synthetic/human
 ;    FEATURE:
 ;      NAME/KEY:  CDS
 ;      LOCATION:  77..409
 ;    FEATURE:
 ;      NAME/KEY:  sig_peptide
 ;      LOCATION:  77..235
 ;    FEATURE:
 ;      NAME/KEY:  mat_peptide
 ;      LOCATION:  236..409
 US-08-321-658B-11


 alignment_scores:
             Quality:  119.00            Length:     63
               Ratio:  3.132               Gaps:      0
  Percent Similarity:  60.317   Percent Identity: 39.683

 alignment_block:
  US-09-215-435-177 x US-08-321-658B-11   ..

  Align seg 1/1  to: US-08-321-658B-11  from: 1  to: 418

       29 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 45
          :::||||||::::::                 |||   |||   |||:::
      212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

       45 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 62
           :::   |||   |||              ||||||||||||||:::
      262 AGATCCTGGAATATGTCGAGGTTATATTACCAGGTATTTTTATAACAATC 311

       62 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 78
          :::|||:::||||||   |||   ::::::||||||   |||||||:::
      312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCCTGGGCAATATG 361

       79 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
          ||||||||||:::         |||   :::       |||
      362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

  seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-135

  seq_documentation_block:
  ; Sequence 135, Application US/08358160
  ; Patent No. 5663143
  ;  GENERAL INFORMATION:
  ;    APPLICANT:  LEY, Arthur C.
  ;    APPLICANT:  LADNER, Robert C.
  ;    APPLICANT:  GUTERMAN, Sonia K.
  ;    APPLICANT:  ROBERTS, Bruce L.
  ;    APPLICANT:  MARKLAND, William
  ;    APPLICANT:  KENT, Rachel B.
  ;    TITLE OF INVENTION:  ENGINEERED HUMAN-DERIVED KUNITZ
  ;    TITLE OF INVENTION:  DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
  ;    NUMBER OF SEQUENCES:  234
  ;    CORRESPONDENCE ADDRESS:
  ;      ADDRESSEE:  BROWDY AND NEIMARK
  ;      STREET:  419 Seventh Street, N.W. Suite 300
```

```
;       CITY:  Washington
;       STATE:  District of Columbia
;       COUNTRY:  USA
;       ZIP:  20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/358,160
;       FILING DATE:  16-DEC-1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/133,031
;       FILING DATE:  13-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/009,319
;       FILING DATE:  26-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/664,989
;       FILING DATE:  01-MAR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/487,063
;       FILING DATE:  02-MAR-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/240,160
;       FILING DATE:  02-SEP-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cooper, Iver P.
;       REGISTRATION NUMBER:  28,005
;       REFERENCE/DOCKET NUMBER:  LEY=1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-628-5197
;       TELEFAX:  202-737-3528
;       TELEX:  248633
;   INFORMATION FOR SEQ ID NO:  135:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  249 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  synthetic DNA fragment
US-08-358-160-135


alignment_scores:
            Quality:  118.00            Length:     70
              Ratio:  2.950               Gaps:      0
 Percent Similarity:  57.143   Percent Identity: 37.143

alignment_block:
 US-09-215-435-177 x US-08-358-160-135   ..

 Align seg 1/1  to: US-08-358-160-135  from: 1  to: 249

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::     ::::::            |||
      10 CTTCTCTTCGCCATTCCTCTGGTGGTACCTTTCTATTCCGGCGCCAAGCC 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         |||   |||   |||:::   :::   |||   |||
      60 TGACTTCTGCTTCCTCGAGGAGGATCCCGGGATTTGCCGCGGTTATATTA 109

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||:::   :::|||:::||||||   |||   :::
     110 CGCGTTATTTCTATAATAACCAGACTAAGCAATGTGAGCGGTTCAAGTAT 159

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||   ||||||:::|||||||||:::         |||   ::
     160 GGTGGTTGCCTAGGTAATATGAACAACTTCGAGACTCTAGAAGAGTGTAA 209

      88 uValAlaCys 91
         :      |||
     210 GAACATATGT 219

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-2

seq_documentation_block:
; Sequence 2, Application US/07828920A
; Patent No. 5312736
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Jesper
;    APPLICANT:  No. 5312736dfang, Ole Juul
;    TITLE OF INVENTION:  Anticoagulant Protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5312736o No. 5312736disk of No. 5312736th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6200
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/828,920A
;      FILING DATE:  19920127
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 4080/89
;      FILING DATE:  18-AUG-1989
;      APPLICATION NUMBER:  WO PCT/DK90/00212
;      FILING DATE:  17)AUG)1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zelson, Steve T.
;      REGISTRATION NUMBER:  30335
;      REFERENCE/DOCKET NUMBER:  3287.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..437
US-07-828-920A-2


alignment_scores:
            Quality:  116.00            Length:     53
              Ratio:  3.515               Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 43.396

alignment_block:
 US-09-215-435-177 x US-07-828-920A-2   ..

 Align seg 1/1  to: US-07-828-920A-2  from: 1  to: 474

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||   |||:::   :::   |||   |||
     234 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 283

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
           ||||||||||||||:::   :::|||:::||||||   |||   ::::
     284 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 333
```

```
     72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
        ::|||||||   |||||||:::||||||||||:::          |||   :::
    334 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 383

     89 ValAlaCys 91
            |||
    384 AACATTTGT 392

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-4

seq_documentation_block:
; Sequence 4, Application US/07828920A
; Patent No. 5312736
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Jesper
;    APPLICANT:  No. 5312736dfang, Ole Juul
;    TITLE OF INVENTION:  Anticoagulant Protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5312736o No. 5312736disk of No. 5312736th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6200
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/828,920A
;      FILING DATE:  19920127
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 4080/89
;      FILING DATE:  18-AUG-1989
;      APPLICATION NUMBER:  WO PCT/DK90/00212
;      FILING DATE:  17)AUG)1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zelson, Steve T.
;      REGISTRATION NUMBER:  30335
;      REFERENCE/DOCKET NUMBER:  3287.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  693 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..656
US-07-828-920A-4


 alignment_scores:
             Quality:  116.00              Length:     53
               Ratio:  3.515                Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 43.396

 alignment_block:
  US-09-215-435-177 x US-07-828-920A-4    ..

  Align seg 1/1  to: US-07-828-920A-4  from: 1  to: 693

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||  |||:::   :::   |||   |||
     453 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 502

     55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
        ||||||||||||||||:::   :::|||:::|||||||   |||   ::::
    503 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 552

     72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
        ::|||||||  |||||||:::||||||||||:::          |||   :::
    553 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 602

     89 ValAlaCys 91
            |||
    603 AACATTTGT 611

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-026-145-1

seq_documentation_block:
; Sequence 1, Application US/08026145
; Patent No. 5378614
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Jens G. Litske
;    APPLICANT:  No. 5378614dfang, Ole Juul
;    TITLE OF INVENTION:  Method for Making TFPI Analogues
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5378614o No. 5378614disk of No. 5378614th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6200
;      CITY:  New York
;      STATE:  N. Y.
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/026,145
;      FILING DATE:  19930302
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/828,920
;      FILING DATE:  27-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DK/90/00212
;      FILING DATE:  17-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 4080/89
;      FILING DATE:  18-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl H.
;      REGISTRATION NUMBER:  34086
;      REFERENCE/DOCKET NUMBER:  3321.214-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  928 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..919
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  8..91
;    FEATURE:
;      NAME/KEY:  mat_peptide
```

```
;      LOCATION:  92..919
US-08-026-145-1


alignment_scores:
            Quality:  116.00              Length:      53
              Ratio:  3.515                 Gaps:       0
 Percent Similarity:  62.264    Percent Identity: 43.396

alignment_block:
 US-09-215-435-177 x US-08-026-145-1   ..

 Align seg 1/1  to: US-08-026-145-1  from: 1  to: 928

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||   |||:::   :::   |||   |||
     374 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 423

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||:::   :::|||:::||||||   |||   ::::
     424 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 473

      72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
         ::||||||   ||||||:::||||||||:::         |||   :::
     474 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 523

      89 ValAlaCys 91
            |||
     524 AACATTTGT 532

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-446-646-8

seq_documentation_block:
; Sequence 8, Application US/08446646
; Patent No. 5726038
;  GENERAL INFORMATION:
;    APPLICANT:  Christiansen, Lars
;    APPLICANT:  Petersen, Jens G.
;    TITLE OF INVENTION:  A DNA Construct Encoding the YAP3 Signal
;    TITLE OF INVENTION:  Peptide
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5726038o No. 5726038disk of No. 5726038th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,646
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3987.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..919
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  8..91
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  92..919
US-08-446-646-8


alignment_scores:
            Quality:  116.00              Length:      53
              Ratio:  3.515                 Gaps:       0
 Percent Similarity:  62.264    Percent Identity: 43.396

alignment_block:
 US-09-215-435-177 x US-08-446-646-8   ..

 Align seg 1/1  to: US-08-446-646-8  from: 1  to: 928

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||   |||:::   :::   |||   |||
     374 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 423

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||:::   :::|||:::||||||   |||   ::::
     424 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 473

      72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
         ::||||||   ||||||:::||||||||:::         |||   :::
     474 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 523

      89 ValAlaCys 91
            |||
     524 AACATTTGT 532

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-7

seq_documentation_block:
; Sequence 7, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
```

; REFERENCE/DOCKET NUMBER: 3693.210-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 945 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 365..538
US-08-321-658B-7

alignment_scores:
Quality: 116.00 Length: 53
Ratio: 3.515 Gaps: 0
Percent Similarity: 62.264 Percent Identity: 43.396

alignment_block:
US-09-215-435-177 x US-08-321-658B-7 ..

Align seg 1/1 to: US-08-321-658B-7 from: 1 to: 945

```
     39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
        |||   |||   |||:::   :::   |||   |||
    371 GATTTCTGCTTTTTGGAAGAAGATCCTGGAATATGTCGAGGTTATATTAC 420

     55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||:::   :::|||:::||||||   |||   ::::
    421 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAACGTTTCAAGTATG 470

     72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
        ::|||||   ||||||:::||||||||||:::         |||   :::
    471 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTGGAAGAATGCAAG 520

     89 ValAlaCys 91
           |||
    521 AACATTTGT 529
```

```
OM of: US-09-215-435-177 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:54 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence          Strd Orig    ZScore  EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799 +  539.00 1232.39  2.7e-59   499
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799 +  539.00 1232.39  2.7e-59   49
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76 +  539.00 1231.85  2.8e-59   526
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303 +  529.00 1208.47  5.7e-58   539
/cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136 +  529.00 1208.47  5.7e-58   539
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148 +  427.00  979.45  3.3e-45   271
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-516-448-1671 +  368.00  840.41  1.8e-37   34
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-833-381-1671 +  368.00  840.41  1.8e-37   349
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-843-620-654 +  368.00  840.41  1.8e-37   349
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-188-162-1870 -  335.00  758.86  6.3e-33   58
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-208-69337 +  251.00  571.90  1.6e-22   2
/cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-131-380-189 +  210.00  480.12  2.1e-17   208
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-499-23002 +  210.00  480.12  2.1e-17   2
/cgn1_7/ptodata/1/pna/US6005_COMB.seq:US-60-055-705-189 +  210.00  480.12  2.1e-17   208
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-6802 +  134.00  297.61  3.1e-07   404
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-960-352-6802 +  134.00  297.61  3.1e-07   404
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-13925 +  134.00  297.24  3.3e-07   41
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-960-352-13925 +  134.00  297.24  3.3e-07   41
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-14462A-23 +  133.00  248.34  0.0002   425
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-01313-23 +  133.00  248.34  0.0002   4252
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-235-810-23 +  133.00  248.34  0.0002   42521
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658-23 +  133.00  248.34  0.0002   42521
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658B-23 +  133.00  248.34  0.0002   4252
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658C-23 +  133.00  248.34  0.0002   4252
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-171-503-23 +  133.00  248.34  0.0002   4252
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6889 +  132.50  293.10  5.5e-07   44
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-6788 +  132.50  291.29  7.0e-07   53
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-142-896-264 -  132.50  275.64  5.2e-06   253
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-6787 +  132.50  275.62  5.2e-06   25
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6888 -  132.50  274.29  6.2e-06   28
/cgn1_7/ptodata/1/pna/US6015_COMB.seq:US-60-150-584-381 +  132.50  262.75  2.7e-05   908
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-237A-859 -  132.50  231.90  0.0014   19
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-849-406-3 +  131.00  294.19  4.8e-07   285
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-24855 +  131.00  272.72  7.6e-06   2
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6885 -  131.00  265.98  1.8e-05   46
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-143-992-158 -  130.50  270.77  9.7e-06   259
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-289-768-38952 +  130.00  286.41  1.3e-06   49
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-331-793-28 +  129.50  286.32  1.3e-06   441
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-235-076-30403 +  129.50  284.81  1.6e-06   51
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-289-768-37096 +  129.50  284.81  1.6e-06   51
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-332-782-30403 +  129.50  284.81  1.6e-06
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-737-223-30403 +  129.50  284.81  1.6e-06
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-203-3184 +  129.50  284.39  1.7e-06
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-190-812 +  129.50  284.37  1.7e-06
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-203-2095 +  129.50  284.37  1.7e-06

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799

seq_documentation_block:
; Sequence 5799, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 5799
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 22..318
;   NAME/KEY: sig_peptide
;   LOCATION: 22..93
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
US-09-834-366-5799


alignment_scores:
            Quality:  539.00              Length:     99
              Ratio:  5.444                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177 x US-09-834-366-5799   ..

 Align seg 1/1  to: US-09-834-366-5799  from: 1  to: 499

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      72 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 121

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     122 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 171

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     172 TATGAAGTWCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 221

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     222 AACTTTTGTCTTCTCCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 271

      84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
         |||||||||||||||||||||||||||||||||||||||||||||||
     272 AAATAGAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACCGAGG 318
```

```
seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799

seq_documentation_block:
; Sequence 5799, Application US/60197873
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
;  FILE REFERENCE:  81.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/197,873
;  CURRENT FILING DATE:  2000-04-18
;  NUMBER OF SEQ ID NOS:  52153
;  SOFTWARE:  Patent.pm
; SEQ ID NO 5799
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 22..318
;   NAME/KEY: sig_peptide
;   LOCATION: 22..93
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
US-60-197-873-5799


alignment_scores:
            Quality:  539.00              Length:     99
              Ratio:  5.444                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177 x US-60-197-873-5799   ..

 Align seq 1/1  to: US-60-197-873-5799  from: 1  to: 499

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      72 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 121

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     122 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 171

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     172 TATGAAGTWCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 221

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     222 AACTTTTGTCTTCTCCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 271

      84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
         |||||||||||||||||||||||||||||||||||||||||||||||
     272 AAATAGAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACCGAGG 318

 seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76

 seq_documentation_block:
 ; Sequence 76, Application US/09215435A
 ; GENERAL INFORMATION:
 ;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
 ;  APPLICANT: Duclert Aymeric,
 ;  APPLICANT: Bougueleret Lydie
 ;  TITLE OF INVENTION: Extended cDNAs
 ;  FILE REFERENCE: GENSET.019A
 ;  CURRENT APPLICATION NUMBER: US/09/215,435A
 ;  CURRENT FILING DATE: 1998-12-17
 ;  EARLIER APPLICATION NUMBER: 60/069,957
 ;  EARLIER FILING DATE: 1997-12-17
 ;  EARLIER APPLICATION NUMBER: 60/074,121
 ;  EARLIER FILING DATE: 1998-2-9
 ;  EARLIER APPLICATION NUMBER: 60/081,563
 ;  EARLIER FILING DATE: 1998-4-13
 ;  EARLIER APPLICATION NUMBER: 60/096,116
 ;  EARLIER FILING DATE: 1998-8-10
 ;  NUMBER OF SEQ ID NOS: 519
 ;  SOFTWARE: Patent.pm
 ; SEQ ID NO 76
 ;   LENGTH: 526
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 22..318
 ;   FEATURE:
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 22..93
 ;   OTHER INFORMATION: Von Heijne matrix
 ;   OTHER INFORMATION: score 4.6
 ;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
 ;   FEATURE:
 ;   NAME/KEY: polyA_signal
 ;   LOCATION: 497..502
 ;   FEATURE:
 ;   NAME/KEY: polyA_site
 ;   LOCATION: 516..526
 US-09-215-435-76


 alignment_scores:
             Quality:  539.00              Length:     99
               Ratio:  5.444                Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-177 x US-09-215-435-76   ..

  Align seq 1/1  to: US-09-215-435-76  from: 1  to: 526

        1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

       17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       72 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 121

       34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      122 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 171

       51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      172 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 221

       67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      222 AACTTTTGTCTTCTCCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 271

       84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
          |||||||||||||||||||||||||||||||||||||||||||||||
      272 AAATAGAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACCGAGG 318

  seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303

  seq_documentation_block:
```

```
; Sequence 303, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 303
;   LENGTH: 539
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 32..328
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 32..103
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 508..513
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 528..539
US-09-215-435-303


alignment_scores:
            Quality:  529.00              Length:     99
              Ratio:  5.343                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.980

alignment_block:
 US-09-215-435-177 x US-09-215-435-303   ..

 Align seg 1/1  to: US-09-215-435-303  from: 1  to: 539

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 131

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     132 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 181

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     182 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 231

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     232 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 281

      84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     282 AAATAGAACGTGAAGTAKCCTGTGTTGCAAAATACAAACCACCGAGG 328

seq_name: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136

seq_documentation_block:
; Sequence 136, Application US/60069957
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  EXTENDED cDNAS FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  381
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/069,957
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.019PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  136:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  32..328
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  508..513
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  528..539
;      IDENTIFICATION METHOD:  blastn
US-60-069-957-136


alignment_scores:
            Quality:  529.00              Length:     99
              Ratio:  5.343                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.980

alignment_block:
 US-09-215-435-177 x US-60-069-957-136   ..

 Align seg 1/1  to: US-60-069-957-136  from: 1  to: 539
```

```
      1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

     17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     82 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 131

     34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    132 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 181

     51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    182 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 231

     67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
        ||||||||||||||:::|||||||||||||||||||||||||||||||||
    232 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAACAACTTCAAGCTTA 281

     84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProProArg 99
        ||||||||||||||||:::|||||||||||||||||||||||||||
    282 AAATAGAACGTGAAGTAKCCTGTGTTGCAAAATACAAACCACCGAGG 328
```

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148

```
seq_documentation_block:
; Sequence 148, Application US/08905279
;  GENERAL INFORMATION:
;    APPLICANT:  Genset SA
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,279
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.011A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  148:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
US-08-905-279-148
```

```
alignment_scores:
            Quality:  427.00              Length:      79
              Ratio:  5.405                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 98.734

alignment_block:
 US-09-215-435-177 x US-08-905-279-148    ..

 Align seg 1/1  to: US-08-905-279-148  from: 1  to: 271

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATAT 131

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     132 GTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGC 181

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     182 TATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCAAAAGATGTGA 231

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 79
         ||||||||||||||||:::|||||||||||||||||||
     232 AACTTTTGTCTTCTCCAGCTGTAATGGCAACCTTAAC 268

 seq_name: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-516-448-1671

 seq_documentation_block:
 ; Sequence 1671, Application US/09516448
 ; GENERAL INFORMATION:
 ;  APPLICANT:  Robison, Keith E.
 ;  TITLE OF INVENTION:  Novel Nucleic Acid and Protein Homologs
 ;  FILE REFERENCE:  5800-119
 ;  CURRENT APPLICATION NUMBER: US/09/516,448
 ;  CURRENT FILING DATE:  2000-02-29
 ;  NUMBER OF SEQ ID NOS:  2050
 ;  SOFTWARE:  FastSEQ for Windows Version 3.0
 ; SEQ ID NO 1671
 ;   LENGTH: 349
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: misc_feature
 ;   LOCATION: (1)...(349)
 ;   OTHER INFORMATION: n = A,T,C or G
 US-09-516-448-1671


 alignment_scores:
             Quality:  368.00              Length:      68
               Ratio:  5.662                 Gaps:       0
  Percent Similarity:  95.588   Percent Identity: 95.588

 alignment_block:
  US-09-215-435-177 x US-09-516-448-1671    ..

  Align seg 1/1  to: US-09-516-448-1671  from: 1  to: 349

       32 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 48
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

       48 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 65
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122
```

```
      65 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 81
         ||||||||||||||   ||||||||||||||||||||||||||||||||
     123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

      82 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 98
         ||||||||||||   ||||||||||||||||||||||||||||||||||
     173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

      98 oArg 99
         ||||
     223 GAGG 226

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-833-381-1671

seq_documentation_block:
; Sequence 1671, Application US/09833381
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: Novel Nucleic Acid and Protein Homologs
;  FILE REFERENCE: 5800-119
;  CURRENT APPLICATION NUMBER: US/09/833,381
;  CURRENT FILING DATE:  2001-04-11
;  PRIOR APPLICATION NUMBER: 09/516,448
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 2050
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1671
;   LENGTH: 349
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(349)
;   OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1671


alignment_scores:
           Quality:  368.00            Length:     68
             Ratio:  5.662               Gaps:      0
 Percent Similarity:  95.588   Percent Identity: 95.588

alignment_block:
 US-09-215-435-177 x US-09-833-381-1671    ..

 Align seg 1/1  to: US-09-833-381-1671  from: 1  to: 349

      32 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

      48 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 65
         ||||||||||||||||||||||||||||||||||||||||||||||||   |
      73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122

      65 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 81
         ||||||||||||||   ||||||||||||||||||||||||||||||||
     123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

      82 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 98
         ||||||||||||   ||||||||||||||||||||||||||||||||||
     173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

      98 oArg 99
         ||||
     223 GAGG 226

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-843-620-654

seq_documentation_block:
; Sequence 654, Application US/09843620
; GENERAL INFORMATION:
;  APPLICANT: Gearing, David P.
;  APPLICANT:  Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.2098-001
;  CURRENT APPLICATION NUMBER: US/09/843,620
;  CURRENT FILING DATE:  2001-04-26
;  PRIOR APPLICATION NUMBER: US 60/200,099
;  PRIOR FILING DATE: 2000-04-27
;  NUMBER OF SEQ ID NOS: 1135
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 654
;   LENGTH: 349
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(349)
;   OTHER INFORMATION: n = A,T,C or G
US-09-843-620-654


alignment_scores:
           Quality:  368.00            Length:     68
             Ratio:  5.662               Gaps:      0
 Percent Similarity:  95.588   Percent Identity: 95.588

alignment_block:
 US-09-215-435-177 x US-09-843-620-654    ..

 Align seg 1/1  to: US-09-843-620-654  from: 1  to: 349

      32 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

      48 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 65
         ||||||||||||||||||||||||||||||||||||||||||||||||   |
      73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122

      65 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 81
         ||||||||||||||   ||||||||||||||||||||||||||||||||
     123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

      82 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 98
         ||||||||||||   ||||||||||||||||||||||||||||||||||
     173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

      98 oArg 99
         ||||
     223 GAGG 226

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-188-162-1870

seq_documentation_block:
; Sequence 1870, Application US/60188162
; GENERAL INFORMATION:
;  APPLICANT:  Beasley, Ellen
;  TITLE OF INVENTION:  ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL000518
;  CURRENT APPLICATION NUMBER: US/60/188,162
;  CURRENT FILING DATE:  2000-05-09
;  NUMBER OF SEQ ID NOS:  5094
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1870
;   LENGTH: 585
;   TYPE: DNA
;   ORGANISM: HUMAN
US-60-188-162-1870


alignment_scores:
```

```
           Quality:  335.00              Length:     60
             Ratio:  5.583                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.333

alignment_block:
 US-09-215-435-177 x US-60-188-162-1870/rev   ..

 Align seg 1/1  to reverse of: US-60-188-162-1870  from: 1  to: 585

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 207

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 157

      72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
         ||:::||||||||||||||||||||||||||||||||||||||||||||||
     156 CCAGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 107

      89 ValAlaCysValAlaLysTyrLysProPro 98
         ||||||||||||||||||||||||||||||
     106 GTAGCCTGTGTTGCAAAATACAAACCACCG 77

seq_name: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-208-69337

seq_documentation_block:
; Sequence 69337, Application US/09540208
; GENERAL INFORMATION:
;  APPLICANT:  Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES OF MALE REPRODUCTIVE TISSUE
;  FILE REFERENCE:  PD-1029 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,208
;  CURRENT FILING DATE:  2000-03-31
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS:  70811
;  SOFTWARE:  PERL Program
; SEQ ID NO 69337
;   LENGTH: 282
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: rat00103589
;   NAME/KEY: unsure
;   LOCATION: 22, 255
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-208-69337


 alignment_scores:
            Quality:  251.00              Length:     71
              Ratio:  4.328                Gaps:      0
 Percent Similarity:  81.690   Percent Identity: 59.155

 alignment_block:
  US-09-215-435-177 x US-09-540-208-69337   ..

  Align seg 1/1  to: US-09-540-208-69337  from: 1  to: 282

       26 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 42
          |||:::|||::|||   :::   |||   :::::::::|||||||||
        1 GTGGAGAAGGTTGCTGCAAGGNTCTGCAAAGAATTCAACGATCCTTGTGC 50

       42 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 59
          |||:::|||:::   ||||||||||||||||   |||:::|||:::||||
       51 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 100
```

```
       59 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 75
          |||||:::||||||||||:::|||:::   ||||||||||||||||||||
      101 ACAACCAAACCTCCAAAAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAAT 150

       76 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysVa 92
          ||||||   |||||||||   ||||||||||:::   :::|||   |||
      151 GGTAACCGTAATAACTTCATGCTTAAAATAGACTGTGACGTAATGTGTAA 200

       92 lAlaLysTyrLys 96
          :::|||||||
      201 TGAAGAATACAAG 213

seq_name: /cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-131-380-189

seq_documentation_block:
; Sequence 189, Application US/09131380
; GENERAL INFORMATION:
;  APPLICANT: Gooding, Douglas H.
;  APPLICANT: Stuve, Laura L.
;  APPLICANT: Stuart, Susan G.
;  APPLICANT: Naughton, Rebecca E.
;  APPLICANT: Garrow, Bonnie L.
;  APPLICANT: Klemm, Juli D.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM RAT FEMUR
;  FILE REFERENCE: PZ-0013 US
;  CURRENT APPLICATION NUMBER: US/09/131,380
;  CURRENT FILING DATE: 1998-08-07
;  EARLIER APPLICATION NUMBER: 60/055,705
;  EARLIER FILING DATE: 1997-08-08
;  EARLIER APPLICATION NUMBER: 60/074,278
;  EARLIER FILING DATE: 1998-02-10
;  NUMBER OF SEQ ID NOS: 2199
;  SOFTWARE: PERL Program
; SEQ ID NO 189
;   LENGTH: 208
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 109, 117, 156, 167
;   OTHER INFORMATION: a or g or c or t, unknown, or other
;   FEATURE: -
;   OTHER INFORMATION: 700292496H1
US-09-131-380-189


 alignment_scores:
            Quality:  210.00              Length:     67
              Ratio:  4.038                Gaps:      1
 Percent Similarity:  77.612   Percent Identity: 55.224

 alignment_block:
  US-09-215-435-177 x US-09-131-380-189   ..

  Align seg 1/1  to: US-09-131-380-189  from: 1  to: 208

       26 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 42
          |||:::|||::|||   :::::::|||   :::::::::|||||||||
        2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

       42 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 59
          |||:::|||:::   ||||||||||||||||   |||:::|||:::||||
       52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

       59 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 75
          |||||   ||||||    :::|||:::   ||||||||||||||||||||
      102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

       75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
          ||||      |||   |||   ||||||||||:::   :::|||   |||
      152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200
```

```
seq_name: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-499-23002

seq_documentation_block:
; Sequence 23002, Application US/09540499
; GENERAL INFORMATION:
;  APPLICANT:  Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES OF MUSCULOSKELETAL SYSTEM TISSUE
;  FILE REFERENCE:  PD-1032 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,499
;  CURRENT FILING DATE:  2000-03-31
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS:  29520
;  SOFTWARE:  PERL Program
; SEQ ID NO 23002
;   LENGTH: 208
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: rat00198915
;   NAME/KEY: unsure
;   LOCATION: 109, 117, 156, 167
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-499-23002


alignment_scores:
            Quality:  210.00            Length:     67
              Ratio:  4.038              Gaps:      1
 Percent Similarity:  77.612   Percent Identity: 55.224

alignment_block:
 US-09-215-435-177 x US-09-540-499-23002   ..

 Align seg 1/1  to: US-09-540-499-23002  from: 1  to: 208

      26 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 42
         |||:::|||:::|||   ::::::|||   ::::::::::|||||||||
       2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

      42 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 59
          |||:::|||:::   |||||||||||||||   |||:::|||:::||||
      52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

      59 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 75
         |||||   ||||||    :::|||:::   ||||||||||||||||||||
     102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         ||||      |||   |||   ||||||||||:::   :::|||   |||
     152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200

seq_name: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:US-60-055-705-189

seq_documentation_block:
; Sequence 189, Application US/60055705
;  GENERAL INFORMATION:
;     APPLICANT:  Gooding, Douglas H.
;     APPLICANT:  Stuart, Susan G.
;     APPLICANT:  Argentine, Charles C.
;     APPLICANT:  Naughton, Rebecca E.
;     TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM RAT
;     TITLE OF INVENTION:  FEMUR
;     NUMBER OF SEQUENCES:  1093
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;       STREET:  3174 PORTER DRIVE
;       CITY:  PALO ALTO
;       STATE:  CALIFORNIA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/60/055,705
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CERRONE, MICHAEL C.
;       REGISTRATION NUMBER:  39,132
;       REFERENCE/DOCKET NUMBER:  PZ-0013 P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 855-0555
;       TELEFAX:  (415) 845-4166
;   INFORMATION FOR SEQ ID NO:  189:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  208 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  700292496H1
US-60-055-705-189


alignment_scores:
            Quality:  210.00            Length:     67
              Ratio:  4.038              Gaps:      1
 Percent Similarity:  77.612   Percent Identity: 55.224

alignment_block:
 US-09-215-435-177 x US-60-055-705-189   ..

 Align seg 1/1  to: US-60-055-705-189  from: 1  to: 208

      26 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 42
         |||:::|||:::|||   :::::::|||   ::::::::::|||||||||
       2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

      42 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 59
          |||:::|||:::   |||||||||||||||   |||:::|||:::||||
      52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

      59 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 75
         |||||   ||||||    :::|||:::   ||||||||||||||||||||
     102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
         ||||      |||   |||   ||||||||||:::   :::|||   |||
     152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-6802

seq_documentation_block:
; Sequence 6802, Application US/09480902
; GENERAL INFORMATION:
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Warren, Wesley C.
;  TITLE OF INVENTION: Nucleic Acid and Other Molecules Associated with Lactation and
;  TITLE OF INVENTION: Muscle and Fat Deposition
;  FILE REFERENCE: 10298/1
;  CURRENT APPLICATION NUMBER: US/09/480,902
;  CURRENT FILING DATE: 2000-01-11
;  EARLIER APPLICATION NUMBER: USSN 60/115,707
;  EARLIER FILING DATE: 1999-01-12
```

```
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6802
;   LENGTH: 404
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 29-LIB34-029-Q1-E1-H1
US-09-480-902-6802


alignment_scores:
            Quality:  134.00              Length:     43
              Ratio:  4.467                 Gaps:      0
 Percent Similarity:  69.767   Percent Identity: 55.814

alignment_block:
 US-09-215-435-177 x US-09-480-902-6802   ..

 Align seg 1/1  to: US-09-480-902-6802  from: 1  to: 404

      39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
         |||   |||:::||||||   :::   |||   |||      :::
     239 GATTCCTGCCAGCTGGACTACTCACAAGGGCCTTGCCTGGGGCTGTTCAA 288

      55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
          ||||||||||||||   ||||||      ||||||||||||:::::::
     289 GAGGTATTTCTATAACGGTACATCCATGGCCTGTGAGACCTTCTTATATG 338

      72 erGlyCysAsnGlyAsnLeuAsnAsnPhe 81
         ::||||||   ||||||   |||||||||
     339 GCGGCTGCATGGGGAATGGCAACAACTTC 367
```

```
OM of: US-09-215-435-177 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:08 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-3 +  131.00  304.97  9.0e-10   28
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-78 +  118.00  274.90  4.3e-08    2
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-84 +  105.00  245.62  1.8e-06    2
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-80 +  100.00  234.16  7.9e-06    1
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-1 +   99.50  229.12  1.5e-05   27
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11942 +   99.50  222.37  3.6e-05
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-149 +   92.50  179.84  0.0084    6
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-9347 +   82.50  183.62  0.0052
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-82 +   81.00  188.21  0.0029    1
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-73 +   79.50  175.95  0.0138    4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11256 +   76.00  168.52  0.0359
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-71 +   75.50  136.34  2.22      8
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-75 +   75.50  136.33  2.22      8
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8534 +   75.00  165.68  0.0516
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-830-446-15 +   68.00  131.52  4.13      23
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8985 -   67.50  147.40  0.5383
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8984 -   67.50  146.49  0.6050
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-871-161-390 +   67.50  143.96  0.8370    5
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-871-161-93 -   65.50  138.40  1.71      64
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-981-353-15 +   65.50  118.99  20.58     43
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-981-353-159 +   65.50  116.79  27.27     5
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3786 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3787 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3788 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3789 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3790 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3791 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3792 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3793 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3794 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3795 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3796 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3797 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3798 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3799 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3800 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3801 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3802 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3803 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3804 -   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3805 +   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3806 +   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3807 +   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3808 +   64.50   91.38  710.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3809 +   64.50   91.38  710.15

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-3

seq_documentation_block:
; Sequence 3, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 285 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-849-406A-3


alignment_scores:
            Quality:  131.00              Length:     85
              Ratio:  2.911                 Gaps:      2
 Percent Similarity:  52.941   Percent Identity: 37.647

alignment_block:
 US-09-215-435-177 x US-08-849-406A-3   ..

 Align seg 1/1  to: US-08-849-406A-3  from: 1  to: 285

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::      ::::::           |||      ::
      10 TTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTGGCGCCAAAGA 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
      60 AGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 109
```

```
     55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
        |||||||||||||||   ||||||      ||||||||||||   :::
    110 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 159

     72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGlu..... 86
        :::|||||||   ||||||   |||||||||      :::   |||
    160 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 209

     87 .........ArgGluVal.....................AlaCysValA 93
                  |||   |||                    :::|||:::|
    210 GCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGAAAGTTGTTTAG 259

     93 laLys 94
        |||||
    260 CAAAA 264

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-78

seq_documentation_block:
; Sequence 78, Application US/08849406A
;         APPLICANT: LEY, Arthur C.
;                    LADNER, Robert C.
;                    GUTERMAN, Sonia K.
;                    ROBERTS, Bruce L.
;                    MARKLAND, William
;                    KENT, Rachel B.
;         TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                             DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;         NUMBER OF SEQUENCES: 86
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;              STREET: 624 Ninth Street, N.W. Suite 300
;              CITY: Washington
;              STATE: District of Columbia
;              COUNTRY: USA
;              ZIP: 20001
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/849,406A
;              FILING DATE: 21-Jul-1999
;              CLASSIFICATION: 530
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: PCT/US95/16349
;              FILING DATE: 15-DEC-1995
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Cooper, Iver P.
;              REGISTRATION NUMBER: 28,005
;              REFERENCE/DOCKET NUMBER: LEY=1A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 202-628-5197
;              TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 78:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 249 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-08-849-406A-78


alignment_scores:
            Quality:  118.00              Length:      70
              Ratio:  2.950                 Gaps:       0
 Percent Similarity:  57.143   Percent Identity:  37.143

alignment_block:
```

```
US-09-215-435-177 x US-08-849-406A-78    ..

Align seg 1/1  to: US-08-849-406A-78  from: 1  to: 249

     22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
        ||||||   ::::::      ::::::          |||
     10 CTTCTCTTCGCCATTCCTCTGGTGGTACCTTTCTATTCCGGCGCCAAGCC 59

     38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
        |||   |||   |||:::   :::   |||   |||
     60 TGACTTCTGCTTCCTCGAGGAGGATCCCGGGATTTGCCGCGGTTATATTA 109

     55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
        ||||||||||||||||:::   :::|||:::||||||   |||   :::
    110 CGCGTTATTTCTATAATAACCAGACTAAGCAATGTGAGCGGTTCAAGTAT 159

     72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
        :::|||||||   ||||||:::||||||||||:::          |||   ::
    160 GGTGGTTGCCTAGGTAATATGAACAACTTCGAGACTCTAGAAGAGTGTAA 209

     88 uValAlaCys 91
        :      |||
    210 GAACATATGT 219

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-84

seq_documentation_block:
; Sequence 84, Application US/08849406A
;         APPLICANT: LEY, Arthur C.
;                    LADNER, Robert C.
;                    GUTERMAN, Sonia K.
;                    ROBERTS, Bruce L.
;                    MARKLAND, William
;                    KENT, Rachel B.
;         TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                             DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;         NUMBER OF SEQUENCES: 86
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;              STREET: 624 Ninth Street, N.W. Suite 300
;              CITY: Washington
;              STATE: District of Columbia
;              COUNTRY: USA
;              ZIP: 20001
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/849,406A
;              FILING DATE: 21-Jul-1999
;              CLASSIFICATION: 530
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: PCT/US95/16349
;              FILING DATE: 15-DEC-1995
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Cooper, Iver P.
;              REGISTRATION NUMBER: 28,005
;              REFERENCE/DOCKET NUMBER: LEY=1A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 202-628-5197
;              TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 84:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 201 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-08-849-406A-84
```

```
alignment_scores:
            Quality:  105.00              Length:     57
              Ratio:  3.387                 Gaps:      0
 Percent Similarity:  54.386   Percent Identity: 36.842

alignment_block:
 US-09-215-435-177 x US-08-849-406A-84   ..

 Align seg 1/1  to: US-08-849-406A-84  from: 1  to: 201

      35 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 51
         |||        |||   |||   |||:::         |||:::|||
       1 GGCGCCAAGCCTGACTTCTGCTTCCTCGAGGAGNRTVVSGGGMNTTGCRT 50

      51 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 68
                    |||   |||||||||:::   :::|||:::|||   :
      51 TGSTNWTTTTMNSCGTTDSTTCTATAATAACCAGGCTAAGCAATGTSWGV 100

      68 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 84
         ::|||   :::::|||||      |||   |||||||||:::
     101 NATTCVHATATGGTGGTTGCVHGGSTAATVBGAACAACTTCGAGACTCTA 150

      85 IleGluArgGluValAlaCys 91
            |||   :::      |||
     151 GAAGAGTGTAAGAACATATGT 171

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-80

seq_documentation_block:
; Sequence 80, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 80:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 189 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-08-849-406A-80


alignment_scores:
            Quality:  100.00              Length:     51
              Ratio:  3.125                 Gaps:      0
 Percent Similarity:  62.745   Percent Identity: 33.333

alignment_block:
 US-09-215-435-177 x US-08-849-406A-80   ..

 Align seg 1/1  to: US-08-849-406A-80  from: 1  to: 189

      41 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 57
         |||   :::      :::   |||   |||      :::      |||::
      22 TGCGCTTTCAAAGCTGATGACGGTCCGTGTAAAGCTATCATGAAACGTTT 71

      57 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 74
         :|||:::|||      ::::::::::||||||   |||::::::::::||||
      72 CTTCTTCAACATTTTCACGCGTCAGTGCGAGGAATTCATTTACGGTGGTT 121

      74 ysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAla 90
         ||:::||||||   |||:::|||:::         |||   :::
     122 GTGAAGGTAACCAGAACCGGTTCGAATCTCTAGAGGAATGTAAGAAGATG 171

      91 Cys 91
         |||
     172 TGC 174

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-1

seq_documentation_block:
; Sequence 1, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 276 base pairs
;             TYPE: nucleic acid
```

```
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-849-406A-1


alignment_scores:
            Quality:   99.50              Length:     77
              Ratio:  2.427                Gaps:      2
 Percent Similarity:  53.247   Percent Identity: 35.065

alignment_block:
 US-09-215-435-177 x US-08-849-406A-1   ..

 Align seg 1/1  to: US-08-849-406A-1  from: 1  to: 276

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||   ::::::      ::::::             |||
      10 TTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTGGCGCCCGTCC 59

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
          |||   |||                |||   |||            :
      60 GGATTTCTGTCTCGAGCCACCATACACTGGGCCCTGCAAAGCGCGCATCA 109

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
         ::|||||||||||||||      :::      |||:::|||||||||:::
     110 TCCGCTATTTCTACAATGCTAAAGCAGGCCTGTGCCAGACCTTTGTATAC 159

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||:::::::::   ||||||||||||         :::   ||
     160 GGTGGTTGCCGTGCTAAGCGTAACAACTTTAAA.........TCGGCCGA 200

      88 u.ValAlaCysValAlaLysTyrLysPro 97
          | :::|||||||||.          |||
     201 AGATTGCATGCGTACCTGCGGTGGCGCCG 229

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11942

seq_documentation_block:
; Sequence 11942, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11942
;   LENGTH: 538
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(538)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-11942


alignment_scores:
            Quality:   99.50              Length:     59
              Ratio:  3.109                Gaps:      2
 Percent Similarity:  54.237   Percent Identity: 38.983

alignment_block:
 US-09-215-435-177 x US-09-922-340-11942   ..
```

```
Align seg 1/1  to: US-09-922-340-11942  from: 1  to: 538

      34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
         |||   ||||||   |||                          |||   ||
     331 TGCGTGGACCTGCCAGACACA.....................GGACTCTG 359

      50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
         |   |||         |||:::::::||||||      |||::::::|||
     360 CAAGGAGAGCATCCCGCGCTGGTACTACAACCCCTTCAGTGAACACTGCG 409

      67 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 83
             |||::::::::::||||||   ||||||   |||||||||
     410 CCCGCTTTACCTATGGTGGTTGTTATGGCAACAAGAACAACTTT..... 453

      84 LysIleGluArgGluValAlaCysVal 92
             |||:::|||      |||:::
     454 ......GAGGAAGAGCAGCAGTGCCTC 474

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-149

seq_documentation_block:
; Sequence 149, Application US/09971392
; GENERAL INFORMATION:
;  APPLICANT: Peterson, David P.
;  APPLICANT:  Pearson, Cecelia I.
;  APPLICANT:  Cocks, Benjamin G.
;  TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
;  FILE REFERENCE: PA-0029 US
;  CURRENT APPLICATION NUMBER: US/09/971,392
;  CURRENT FILING DATE:  2001-10-03
;  PRIOR APPLICATION NUMBER: 60/237,652
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 260
;  SOFTWARE: PERL Program
; SEQ ID NO 149
;   LENGTH: 6777
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Template ID: 235164.25
US-09-971-392-149


alignment_scores:
            Quality:   92.50              Length:     68
              Ratio:  2.312                Gaps:      2
 Percent Similarity:  58.824   Percent Identity: 26.471

alignment_block:
 US-09-215-435-177 x US-09-971-392-149   ..

 Align seg 1/1  to: US-09-971-392-149  from: 1  to: 6777

      26 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 42
         |||:::::::::::::::::::::|||:::::      :::
     982 GTGGAAGAGGTGGTTCGAGAGGTGTGCTCTGAACAAGCCGAGACG..... 1026

      42 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 59
                    |||   |||      :::      |||:::::::
    1027 ..............GGGCCGTGCCGAGCAATGATCTCCCGCTGGTACT 1060

      59 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 75
         :::::   |||:::   ::::||      |||   ::::::||||||:::
    1061 TTGATGTGACTGAAGGGAAGTGTGCCCCATTCTTTTACGGCGGATGTGGC 1110

      76 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysVa 92
         ||||||   |||||||||           :::   |||      |||::
    1111 GGCAACCGGAACAACTTT...........GACACAGAAGAGTACTGCAT 1148

      92 lAla 93
         :|||
    1149 GGCC 1152
```

```
seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-9347

seq_documentation_block:
; Sequence 9347, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9347
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(426)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-9347


alignment_scores:
           Quality:   82.50            Length:     49
             Ratio:  3.056              Gaps:      2
 Percent Similarity:  55.102   Percent Identity: 40.816

alignment_block:
 US-09-215-435-177 x US-09-922-340-9347   ..

 Align seg 1/1  to: US-09-922-340-9347  from: 1  to: 426

      34 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 50
         |||   ||||||   |||                        |||   ||
     234 TGCGTGGACCTGCCAGACACA.....................GGACTCTG 262

      50 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 67
         |   |||         |||::::::||||||      |||:::::::|||
     263 CAAGGAGAGCATCCCGCGCTGGTACTACAACCCCTTCAGCGAACACTGCG 312

      67 luThrPheValPheSerGlyCysAsnGlyAsn.LeuAsnAsnPhe 81
            |||::::::::::||||||   ||||||    |||||||||
     313 CCCGCTTTACCTATGGTGGTTGTTACGGNAACAANGAACAACTTT 357

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-82

seq_documentation_block:
; Sequence 82, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 82:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 189 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-08-849-406A-82


alignment_scores:
           Quality:   81.00            Length:     51
             Ratio:  3.000              Gaps:      0
 Percent Similarity:  52.941   Percent Identity: 29.412

alignment_block:
 US-09-215-435-177 x US-08-849-406A-82   ..

 Align seg 1/1  to: US-08-849-406A-82  from: 1  to: 189

      41 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 57
         |||   :::              |||:::|||             |||
      22 TGCGCTTTCAAAGCTNRTRVSGGTCNTTGTRTTGSTNTCTTCMNSCGTTD 71

      57 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 74
          |||:::|||      ::::::::::|||   :::|||   ::::::||||
      72 STTCTTCAACATTTTCACGCGTCAGTGCSWGVHATTCVHATACGGTGGTT 121

      74 ysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAla 90
         ||      |||   |||:::|||:::         |||   :::
     122 GTVHGGSTAACSRGAACCGGTTCGAATCTCTAGAGGAATGTAAGAAGATG 171

      91 Cys 91
         |||
     172 TGC 174

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-73

seq_documentation_block:
; Sequence 73, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
```

```
;          ZIP: 20001
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/849,406A
;            FILING DATE: 21-Jul-1999
;            CLASSIFICATION: 530
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/US95/16349
;            FILING DATE: 15-DEC-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Cooper, Iver P.
;            REGISTRATION NUMBER: 28,005
;            REFERENCE/DOCKET NUMBER: LEY=1A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 202-628-5197
;            TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 73:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 444 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-08-849-406A-73


alignment_scores:
            Quality:   79.50              Length:    113
              Ratio:   1.472                Gaps:      4
 Percent Similarity:   47.788   Percent Identity: 28.319

alignment_block:
 US-09-215-435-177 x US-08-849-406A-73    ..

 Align seg 1/1  to: US-08-849-406A-73  from: 1  to: 444

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhe..............  12
         :::|||:::      |||   ||||||||||||:::
      86 CTGAAGACGAGACTGCTCAAATTCCTGCTGAGGCTGTCATCGGTTACTCT 135

      13 .....................PheIlePheCysSerLeuAsnThrLeuL  22
                               :::|||||||        |||||||
     136 GACTTGGAAGGTGACTTCGACGTCGCTGTTTTGCCATTCTCTAACTCTAC 185

      22 euLeu..Gly.................GlyValAsnLysIleAlaGluL  32
         ||     |||                 :::   :::||||||   |
     186 TAACAACGGTTTGTTGTTCATCAACACTACCATCGCTTCTATCGCTGCTA 235

      32 ysIleCysGly.........AspLeuLysAspProCysLysLeuAspMet  45
         ||      |||         |||   ::::::   |||:::|||   :::
     236 AGGAGGAAGGTGTTTCCTTGGACAAGAGAGAGGCTTGTAACTTGCCAATC 285

      46 AsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgTh  62
            |||   |||                |||:::   ::::::   ::
     286 GTCAGAGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGT 335

      62 rSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuA  79
         ::::   :::|||      |||   ::::::||||||:::||||||   |
     336 TAAGGGTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTA 385

      79 snAsnPheLysLeuLysIleGluArgGluValAlaCys  91
         ||:::|||      :::   |||   :::       |||
     386 ACAAGTTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 423

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11256

 seq_documentation_block:
 ; Sequence 11256, Application US/09922340
```

```
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11256
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-340-11256


alignment_scores:
            Quality:   76.00              Length:     34
              Ratio:   3.455                Gaps:      1
 Percent Similarity:   64.706   Percent Identity: 44.118

alignment_block:
 US-09-215-435-177 x US-09-922-340-11256    ..

 Align seg 1/1  to: US-09-922-340-11256  from: 1  to: 401

      59 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs  75
         ||||||      |||::::::|||       |||::::::::::||||||
       2 TACAACCCCTTCAGTGAACACTGCGCCCGCTTTACCTATGGTGGTTGTTA  51

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysV  92
          ||||||   |||||||||           |||:::|||      |||:
      52 TGGCAACAAGAACAACTTT.....GAGGAAGAGCAGCAGTGCC  89

      92 al 92
          ::
      90 TC 91

 seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-71

 seq_documentation_block:
 ; Sequence 71, Application US/08849406A
 ;       APPLICANT: LEY, Arthur C.
 ;                  LADNER, Robert C.
 ;                  GUTERMAN, Sonia K.
 ;                  ROBERTS, Bruce L.
 ;                  MARKLAND, William
 ;                  KENT, Rachel B.
 ;       TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
 ;                           DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
 ;       NUMBER OF SEQUENCES: 86
 ;       CORRESPONDENCE ADDRESS:
 ;            ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
 ;            STREET: 624 Ninth Street, N.W. Suite 300
 ;            CITY: Washington
 ;            STATE: District of Columbia
 ;            COUNTRY: USA
 ;            ZIP: 20001
 ;       COMPUTER READABLE FORM:
 ;            MEDIUM TYPE: Floppy disk
 ;            COMPUTER: IBM PC compatible
 ;            OPERATING SYSTEM: PC-DOS/MS-DOS
 ;            SOFTWARE: PatentIn Release #1.0, Version #1.30
 ;       CURRENT APPLICATION DATA:
 ;            APPLICATION NUMBER: US/08/849,406A
 ;            FILING DATE: 21-Jul-1999
 ;            CLASSIFICATION: 530
 ;        PRIOR APPLICATION DATA:
 ;            APPLICATION NUMBER: PCT/US95/16349
```

```
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8584 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-08-849-406A-71


alignment_scores:
            Quality:   75.50              Length:    113
              Ratio:  1.425                Gaps:      4
 Percent Similarity:  46.903   Percent Identity: 28.319

alignment_block:
 US-09-215-435-177 x US-08-849-406A-71   ..

 Align seg 1/1  to: US-08-849-406A-71  from: 1  to: 8584

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhe..............  12
         :::|||:::     |||   ||||||||||||:::
    1030 CTGAAGACGAGACTGCTCAAATTCCTGCTGAGGCTGTCATCGGTTACTCT 1079

      13 .....................PheIlePheCysSerLeuAsnThrLeuL  22
                              :::|||||||         |||||||
    1080 GACTTGGAAGGTGACTTCGACGTCGCTGTTTTGCCATTCTCTAACTCTAC 1129

      22 euLeu..Gly..................GlyValAsnLysIleAlaGluL  32
         ||     |||                    :::   :::||||||   |
    1130 TAACAACGGTTTGTTGTTCATCAACACTACCATCGCTTCTATCGCTGCTA 1179

      32 ysIleCysGly.........AspLeuLysAspProCysLysLeuAspMet  45
         ||      |||         |||   :::      |||:::|||   :::
    1180 AGGAGGAAGGTGTTTCCTTGGACAAGAGAGCTGCTTGTAACTTGCCAATC 1229

      46 AsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgTh  62
            |||   |||              |||:::   ::::::   ::
    1230 GTCAGAGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGT 1279

      62 rSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuA  79
         ::::   :::|||      |||   :::::::||||||:::||||||   |
    1280 TAAGGGTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTA 1329

      79 snAsnPheLysLeuLysIleGluArgGluValAlaCys  91
         ||:::|||      :::   |||   :::      |||
    1330 ACAAGTTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 1367

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-75

seq_documentation_block:
; Sequence 75, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 75:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8590 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-08-849-406A-75


alignment_scores:
            Quality:   75.50              Length:    113
              Ratio:  1.425                Gaps:      4
 Percent Similarity:  46.903   Percent Identity: 28.319

alignment_block:
 US-09-215-435-177 x US-08-849-406A-75   ..

 Align seg 1/1  to: US-08-849-406A-75  from: 1  to: 8590

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhe..............  12
         :::|||:::     |||   ||||||||||||:::
    1030 CTGAAGACGAGACTGCTCAAATTCCTGCTGAGGCTGTCATCGGTTACTCT 1079

      13 .....................PheIlePheCysSerLeuAsnThrLeuL  22
                              :::|||||||         |||||||
    1080 GACTTGGAAGGTGACTTCGACGTCGCTGTTTTGCCATTCTCTAACTCTAC 1129

      22 euLeu..Gly..................GlyValAsnLysIleAlaGluL  32
         ||     |||                    :::   :::||||||   |
    1130 TAACAACGGTTTGTTGTTCATCAACACTACCATCGCTTCTATCGCTGCTA 1179

      32 ysIleCysGly.........AspLeuLysAspProCysLysLeuAspMet  45
         ||      |||         |||   :::      |||:::|||   :::
    1180 AGGAGGAAGGTGTTTCCTTGGACAAGAGAGCTGCTTGTAACTTGCCAATC 1229

      46 AsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgTh  62
            |||   |||              |||:::   ::::::   ::
    1230 GTCAGAGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGT 1279

      62 rSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuA  79
         ::::   :::|||      |||   :::::::||||||:::||||||   |
    1280 TAAGGGTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTA 1329

      79 snAsnPheLysLeuLysIleGluArgGluValAlaCys  91
         ||:::|||      :::   |||   :::      |||
    1330 ACAAGTTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 1367

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8534
```

```
seq_documentation_block:
; Sequence 8534, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8534
;   LENGTH: 418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(418)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-8534


alignment_scores:
             Quality:   75.00             Length:      33
               Ratio:  3.409               Gaps:       0
 Percent Similarity:  66.667   Percent Identity: 42.424

alignment_block:
 US-09-215-435-177 x US-09-922-340-8534   ..

 Align seg 1/1  to: US-09-922-340-8534  from: 1  to: 418

      59 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs 75
         ||||||      |||:::::::|||      |||::::::::::||||||
          2 TACAACCCCTTCAGTGAACACTGCGCCCGCTTTACCTATGGGGGTTGTTA 51

      75 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 91
          ||||||   ||||||||||:::   |||   :::         :::|||
      52 TGGCAACAAGAACAACTTTGAGGAAAAGCAGCAGTGCCTCGAGTCTTGT 100

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-830-446-15

 seq_documentation_block:
 ; Sequence 15, Application US/09830446
 ; GENERAL INFORMATION:
 ;  APPLICANT: MURDIN, Andrew D.; OOMEN, Raymond P; and WANG, Joe
 ;  TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
 ;  FILE REFERENCE: 77813-51
 ;  CURRENT APPLICATION NUMBER: US/09/830,446
 ;  CURRENT FILING DATE:  2001-09-21
 ;  PRIOR APPLICATION NUMBER: US 60/106034
 ;  PRIOR FILING DATE: 1998-10-28
 ;  PRIOR APPLICATION NUMBER: US 60/106039
 ;  PRIOR FILING DATE: 1998-10-28
 ;  PRIOR APPLICATION NUMBER: US 60/10604
 ;  PRIOR FILING DATE: 1998-10-28
 ;  PRIOR APPLICATION NUMBER: US 60/106044
 ;  PRIOR FILING DATE: 1998-10-28
 ;  PRIOR APPLICATION NUMBER: US 60/106072
 ;  PRIOR FILING DATE: 1998-10-29
 ;  PRIOR APPLICATION NUMBER: US 60/106073
 ;  PRIOR FILING DATE: 1998-10-29
 ;  PRIOR APPLICATION NUMBER: US 60/106074
 ;  PRIOR FILING DATE: 1998-10-29
 ;  PRIOR APPLICATION NUMBER: US 60/106087
 ;  PRIOR FILING DATE: 1998-10-29
 ;  PRIOR APPLICATION NUMBER: US 60/106587
 ;  PRIOR FILING DATE: 1998-11-02
 ;  PRIOR APPLICATION NUMBER: US 60/106588
 ;  PRIOR FILING DATE: 1998-11-02
 ;  PRIOR APPLICATION NUMBER: US 60/107089
 ;  PRIOR FILING DATE: 1998-11-02
 ;  PRIOR APPLICATION NUMBER: US 60/107034
 ;  PRIOR FILING DATE: 1998-11-02
 ;  PRIOR APPLICATION NUMBER: US 60/107035
 ;  PRIOR FILING DATE: 1998-11-02
 ;  NUMBER OF SEQ ID NOS: 45
 ;  SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ;   LENGTH: 2300
 ;   TYPE: DNA
 ;   ORGANISM: Chlamydia pneumoniae
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: (101)..(2194)
 US-09-830-446-15


 alignment_scores:
              Quality:   68.00             Length:      58
                Ratio:  2.061               Gaps:       2
  Percent Similarity:  56.897   Percent Identity: 32.759

 alignment_block:
  US-09-215-435-177 x US-09-830-446-15   ..

  Align seg 1/1  to: US-09-830-446-15  from: 1  to: 2300

       47 PheGly.........SerCysTyrGluValHisPheArgTyrPheTyr.. 59
          ||||||         :::|||   :::   ::::::::||||||   :::
     1522 TTTGGGAGGCTTATTAAGTGCCTTAGATCAAGATGGAGATACTGTCATTG 1571

       60 .............AsnArgThrSerLysArgCysGluThrPheValPheS 72
                      |||:::   |||      :::   :::||||||||||
     1572 TCTTGAATCCTAGAATCATGGCTCAGGATACGCAACAAGCTTCGTTTTTT 1621

       72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
             :::|||||||   |||   |||:::::::::         |||:::
     1622 GTAGGGCAAACGGTCCCTTACCAAACTATCAAATACTATATCCAAGAAAC 1671

       89 ValAlaCysValAlaLysTyrLys 96
             |||   ||||||||||:::
     1672 AGGAACTGTAACGCAAAATATCGA 1695
```

```
OM of: US-09-215-435-177 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:29 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177
Query length: 99
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est45:AW303354      -  433.00 1006.35  6.2e-47   443   ! AW303354 xv18b01.x1 Soares_NFL_
gb_htc:AK020315        +  324.00  746.24  1.9e-32   611   ! AK020315 Mus musculus adult mal
gb_gss32:AZ763929      -  235.00  537.19  8.4e-21   570   ! AZ763929 1M0559C11R Mouse 10kb
gb_est18:AI256949      +  132.50  295.94  2.3e-07   552   ! AI256949 LP04037.5prime LP Dros
gb_est15:AI015730      -  132.00  295.83  2.3e-07   496   ! AI015730 ot71b08.s1 Soares_tota
gb_est109:T74655       +  129.50  293.43  3.2e-07   350   ! T74655 yc57e11.r1 Stratagene li
gb_est7:AA424642       +  129.50  291.77  3.9e-07   413   ! AA424642 zv91d02.r1 Soares_NhHM
gb_est14:AF063548      +  129.50  289.64  5.2e-07   511   ! AF063548 AF063548 Homo sapiens
gb_est16:AI110674      +  129.50  289.64  5.2e-07   511   ! AI110674 HA0104 Human fetal liv
gb_est17:AI207711      +  129.50  289.45  5.3e-07   521   ! AI207711 HA3093 Human fetal liv
gb_est17:AI207590      -  129.50  288.53  5.9e-07   571   ! AI207590 HA2954 Human fetal liv
gb_est40:AV658823      +  129.50  287.61  6.7e-07   626   ! AV658823 AV658823 GLC Homo sapi
gb_est25:AI807643      -  129.50  287.32  6.9e-07   644   ! AI807643 wf49d04.x1 Soares_NFL_
gb_est15:AI065076      +  129.50  287.05  7.2e-07   662   ! AI065076 HA0910 Human fetal liv
gb_est40:AV651758      +  129.50  286.91  7.3e-07   671   ! AV651758 AV651758 GLC Homo sapi
gb_est43:AW173259      -  129.50  286.85  7.4e-07   675   ! AW173259 xj85c08.x1 Soares_NFL_
gb_est40:AV700236      -  129.50  286.68  7.5e-07   687   ! AV700236 AV700236 GKC Homo sapi
gb_est102:BG562602     +  129.50  286.53  7.7e-07   697   ! BG562602 602581372F1 NIH_MGC_76
gb_est102:BG564407     +  129.50  284.88  9.5e-07   822   ! BG564407 602584345F1 NIH_MGC_76
gb_est29:AL563977      -  129.50  284.63  9.8e-07   843   ! AL563977 AL563977 LTI_NFL001_NE
gb_est104:D83828       +  129.00  287.69  6.6e-07   552   ! D83828 HUMHBC4837 Human pancrea
gb_est40:AV657144      +  128.50  291.63  4.0e-07   331   ! AV657144 AV657144 GLC Homo sapi
gb_est40:AV658016      +  128.50  291.04  4.3e-07   351   ! AV658016 AV658016 GLC Homo sapi
gb_est40:AV659081      +  128.50  290.79  4.4e-07   360   ! AV659081 AV659081 GLC Homo sapi
gb_est40:AV682237      +  128.50  290.73  4.5e-07   362   ! AV682237 AV682237 GKB Homo sapi
gb_est40:AV654397      +  128.50  289.81  5.0e-07   397   ! AV654397 AV654397 GLC Homo sapi
gb_est40:AV645675      +  128.50  289.51  5.2e-07   409   ! AV645675 AV645675 GLC Homo sapi
gb_est105:H93888       +  128.50  289.08  5.5e-07   427   ! H93888 yv08g08.r1 Soares fetal
gb_est40:AV720273      -  128.50  287.11  7.1e-07   520   ! AV720273 AV720273 GLC Homo sapi
gb_est40:AV683948      +  128.50  285.58  8.7e-07   606   ! AV683948 AV683948 GKC Homo sapi
gb_est40:AV654309      +  128.50  285.56  8.7e-07   607   ! AV654309 AV654309 GLC Homo sapi
gb_est40:AV655831      +  128.50  285.49  8.8e-07   611   ! AV655831 AV655831 GLC Homo sapi
gb_est40:AV645552      +  128.50  284.75  9.6e-07   658   ! AV645552 AV645552 GLC Homo sapi
gb_est40:AV683159      +  128.50  283.37  1.2e-06   755   ! AV683159 AV683159 GKC Homo sapi
gb_est105:H72811       +  128.00  288.51  6.0e-07   402   ! H72811 yu07f10.r1 Soares fetal
gb_est95:BF990562      +  128.00  287.40  6.9e-07   449   ! BF990562 RC6-GN0071-271000-013-
gb_est42:AW111218      -  128.00  283.65  1.1e-06   653   ! AW111218 MT15218 mouse liver, d
gb_est15:AI065103      +  127.00  283.85  1.1e-06   506   ! AI065103 HA0939 Human fetal liv
gb_est19:AI333042      -  127.00  283.48  1.1e-06   525   ! AI333042 qq31h04.x1 Soares_NhHM
gb_est19:AI336247      -  127.00  282.43  1.3e-06   583   ! AI336247 qt45b07.x1 Soares_feta
gb_est40:AV690405      +  127.00  281.38  1.5e-06   647   ! AV690405 AV690405 GKC Homo sapi
gb_est29:AL551326      +  127.00  277.49  2.4e-06   955   ! AL551326 AL551326 LTI_NFL006
gb_est42:AW111217      -  126.50  281.69  1.4e-06   558   ! AW111217 MT2416 mouse liver,
gb_est95:BG008148      +  126.50  281.21  1.5e-06   585   ! BG008148 QV4-GN0314-291100-6
gb_est17:AI246161      -  126.00  281.87  1.4e-06   487   ! AI246161 qi29b06.x1 Soares_N

seq_name: gb_est45:AW303354

seq_documentation_block:
LOCUS       AW303354      443 bp    mRNA            EST       18-JAN-2000
DEFINITION  xv18b01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2813449 3' similar to SW:IBP_CARCR P00993 CHELONIANIN ;, mRNA
            sequence.
ACCESSION   AW303354
VERSION     AW303354.1  GI:6713034
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2813449"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      137 a     84 c     77 g    145 t
ORIGIN


alignment_scores:
           Quality:  433.00              Length:     78
             Ratio:  5.551                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177 x AW303354/rev   ..

 Align seg 1/1  to reverse of: AW303354  from: 1  to: 443

      22 LeuLeuGlyGlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TTATTGGGTGGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAA 392

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 AGATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACT 342

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
         |||||||||||||||||||||||||||||||||||||||||||||||||
     341 TTAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTC 292
```

```
     72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    291 TCCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGA 242

     88 uValAlaCysValAlaLysTyrLysProProArg 99
        |||||||||||||||||||||||||||||||
    241 AGTAGCCTGTGTTGCAAAATACAAACCACCGAGG 208

seq_name: gb_htc:AK020315

seq_documentation_block:
LOCUS       AK020315      611 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus adult male epididymis cDNA, RIKEN full-length enriched
            library, clone:9230105I15, full insert sequence.
ACCESSION   AK020315
VERSION     AK020315.1  GI:12860868
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male epididymis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:9230105I15.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 611)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
  COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 185.2. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES                Location/Qualifiers
     source             1. .611
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /db_xref="MGD:MGI:1912158"
                        /db_xref="MGD:MGI:1925489"
                        /clone="9230105I15"
                        /sex="male"
                        /tissue_type="epididymis"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="adult"
     CDS                24. .503
                        /note="putative"
                        /codon_start=1
                        /protein_id="BAB32069.1"
                        /db_xref="GI:12860869"
                        /translation="MKPTKLGFLLGLSLLCSLSPPVLSGVERLANYLCKDYNDPCLLD
                        VEPGSCYEVHFRFFYNQTAKQCQIFLFTGCNGNLNNFKLKIDCDVTCHETYKNPPLPS
                        GDKRKRSLRALATKNLATDWRLTTLQPGRLRQTERKETLLQRARRESRRQPLARHHT"
     polyA_signal       591. .596
                        /note="putative"
     polyA_site         611
                        /note="putative"
BASE COUNT      174 a    152 c    145 g    140 t
ORIGIN


alignment_scores:
            Quality:  324.00              Length:     98
              Ratio:  3.951                 Gaps:      0
 Percent Similarity:  83.673   Percent Identity: 58.163

alignment_block:
 US-09-215-435-177 x AK020315   ..

 Align seg 1/1  to: AK020315  from: 1  to: 611

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||   :::||||||||||||||||||   :::   ::::::|||||
      24 ATGAAGCCCACTAAGCTGGGATTTCTCCTGGGACTCTCCCTCCTCTGCTC 73

      17 rLeuAsnThrLeuLeuLeuGlyGlyValAsnLysIleAlaGluLysIleC 34
         ||||:::      :::|||:::||||||||:::::::::|||:::   :::|
      74 ACTCAGTCCTCCAGTACTGAGCGGTGTTGAGAGGCTAGCTAACTATCTCT 123

      34 ysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCys 50
         ||   |||   :::||||||||||   ||||||::::::   |||||||||
     124 GCAAAGACTACAACGATCCCTGTCTGTTGGATGTGGAGCCTGGCAGCTGC 173

      51 TyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGl 67
         ||||||||||||||||||:::||||||||::|||:::|||:::||||::
     174 TATGAAGTACATTTTAGATTTTTCTACAACCAAACTGCCAAACAGTGTCA 223

      67 uThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuL 84
         :   |||:::|||:::||||||||||||||||||||||||||||||||||
     224 GATTTTTCTGTTCACTGGCTGCAATGGCAACCTTAATAACTTCAAGCTGA 273
```

```
     84 ysIleGluArgGluValAlaCysValAlaLysTyrLysProPro 98
        |||||:::   :::|||:::|||         ||||||   |||
    274 AAATAGACTGTGATGTAACTTGTCACGAAACATACAAAAACCCT 317

seq_name: gb_gss32:AZ763929

seq_documentation_block:
LOCUS       AZ763929      570 bp    DNA             GSS       16-FEB-2001
DEFINITION  1M0559C11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0559C11 R, DNA sequence.
ACCESSION   AZ763929
VERSION     AZ763929.1  GI:12878349
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0559  row: C  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0559C11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      179 a    123 c    126 g    142 t
ORIGIN


alignment_scores:
            Quality:  235.00            Length:     60
            Ratio:  4.519               Gaps:       0
 Percent Similarity:  86.667   Percent Identity: 65.000

alignment_block:
 US-09-215-435-177 x AZ763929/rev   ..

 Align seg 1/1  to reverse of: AZ763929  from: 1  to: 570

     39 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 55
        |||||||||   ||||||::::::   ||||||||||||||||||||||
    208 GATCCCTGTCTGTTGGATGTGGAGCCTGGCAGCTGCTATGAAGTACATTT 159

     55 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 72
        ||||:::|||||||||:::|||:::|||:::|||:::   |||:::|||:
    158 TAGATTTTTCTACAACCAAACTGCCAAACAGTGTCAGATTTTTCTGTTCA 109

     72 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 88
        ::|||||||||||||||:::||||||||||||||||||||:::   :::
    108 CTGGCTGCAATGGCAACTTTAATAACTTCAAGCTGAAAATAGACTGTGAT 59

     89 ValAlaCysValAlaLysTyrLysProPro 98
        |||:::|||         ||||||   |||
     58 GTAACTTGTCACGAAACATACAAAAACCCT 29

seq_name: gb_est18:AI256949

seq_documentation_block:
LOCUS       AI256949      552 bp    mRNA            EST       17-NOV-1998
DEFINITION  LP04037.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP04037 5prime, mRNA sequence.
ACCESSION   AI256949
VERSION     AI256949.1  GI:3864474
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Sequence is the complete cDNA Insert. hit genomic sequence DS00216;
            hit genomic sequence DS02911; hit genomic sequence DS03364
            Plate: 40  row: D  column: 1
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LP04037"
                     /clone_lib="LP Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /lab_host="DH5-alpha"
                     /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                     Site_2: Xho1; Sized fractionated cDNAs were directly
                     ligated into pOT2. Plasmid cDNA library.
                     "
BASE COUNT      154 a    124 c    122 g    152 t
ORIGIN


alignment_scores:
            Quality:  132.50            Length:    101
```

```
            Ratio:  2.366              Gaps:      4
Percent Similarity:  55.446   Percent Identity: 31.683

alignment_block:
US-09-215-435-177 x AI256949   ..

Align seq 1/1  to: AI256949  from: 1  to: 552

     12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeuGlyGlyVal..... 26
        |||::::::|||:::::::::      ||||||   :::|||:::
     31 TTCATAGTATTCGCGGCTGTT...CTGCTCCTGGCAGCCGGAATTCGTGC 77

     27 .........AsnLysIleAlaGluLys.............IleCysG 35
                 :::   ::::::||||||               ::::::
     78 AGCTCCCTCGGATGTCGTCGTGGAGAAGGCGGAGGAGCCCGCGGTGGCCA 127

     35 lyAspLeuLysAsp.....................ProCysLysLeu 43
           |||   |||                        |||
    128 AACCCCTGCCTGATGCCAGTGAGCTCACTGTTCCGGAGGACTGTCATCAG 177

     44 AspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAs 60
           :::   |||   |||:::   :::   :::|||||||   |||||
    178 CCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTATCGCTACGCCTACAA 227

     60 nArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyA 77
        |      ::::::   ||||||   ||||||::::::||||||   ||||
    228 CGTGGATACACAATCCTGCGAGGAGTTCGTTTACGGTGGATGTGCCGGCA 277

     77 snLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysValAla 93
        ||   ||||||||||:::   |||   :::   |||   ||||||:::::
    278 ACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAACAGGCATGTTTGGTA 327

     94 Lys 94
        |||
    328 AAG 330

seq_name: gb_est15:AI015730

seq_documentation_block:
LOCUS       AI015730      496 bp    mRNA            EST       27-AUG-1998
DEFINITION  ot71b08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1622199 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN
            );; mRNA sequence.
ACCESSION   AI015730
VERSION     AI015730.1  GI:3230066
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2157   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1622199"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      105 a    145 c    128 g    118 t
ORIGIN


alignment_scores:
            Quality:  132.00            Length:     54
              Ratio:  3.882               Gaps:      0
Percent Similarity:  62.963   Percent Identity: 48.148

alignment_block:
US-09-215-435-177 x AI015730/rev   ..

Align seq 1/1  to reverse of: AI015730  from: 1  to: 496

     38 LysAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHi 54
        :::|||   |||:::|||      :::   |||||||||      :::
    493 GAAGATTCCTGCCAGCTGGGCTACTCGGCCGGTTCCTGCATGGGAATGAC 444

     54 sPheArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValP 71
           ||||||||||||||||   ||||||      ||||||||||||   :
    443 CAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGT 394

     71 heSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArg 87
        :::::||||||   ||||||   |||||||||      :::   |||
    393 ACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGT 344

     88 GluValAlaCys 91
           :::|||
    343 CTGCAGACCTGC 332

seq_name: gb_est109:T74655

seq_documentation_block:
LOCUS       T74655        350 bp    mRNA            EST       02-MAR-1995
DEFINITION  yc57e11.r1 Stratagene liver (#937224) Homo sapiens cDNA clone
            IMAGE:84812 5' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);;
            mRNA sequence.
ACCESSION   T74655
VERSION     T74655.1  GI:691330
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 62
            High qality sequence stops: 281 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 62   Std Error: 0.00
            Seq primer: M13RP1
```

```
           High quality sequence stop: 281.
FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Homo sapiens"
                     /db_xref="GDB:501869"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:84812"
                     /clone_lib="Stratagene liver (#937224)"
                     /sex="male"
                     /dev_stage="49 years old"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Hepatectomy from normal male caucasian. Average insert
                     size: 1.1 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       78 a     92 c    103 g     76 t      1 others
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x T74655   ..

 Align seg 1/1  to: T74655  from: 1  to: 350

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||::::::::::::::|||              ::
       2 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 36

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
      37 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 86

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||||   :::
      87 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 136

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     137 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 186

      88 uValAlaCys 91
            :::|||
     187 GCAGACCTGC 196

seq_name: gb_est7:AA424642

seq_documentation_block:
LOCUS        AA424642      413 bp    mRNA            EST       16-OCT-1997
DEFINITION   zv91d02.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:767139 5'
             similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);, mRNA sequence.
ACCESSION    AA424642
VERSION      AA424642.1  GI:2103595
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 413)
  AUTHORS    Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
             Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
             , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
             ,T., Waterston,R. and Wilson,R.
  TITLE      WashU-Merck EST Project 1997
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:767139"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       93 a    114 c    126 g     80 t
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AA424642   ..

 Align seg 1/1  to: AA424642  from: 1  to: 413

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||::::::::::::::|||              ::
      88 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 122

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     123 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 172

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||   ||||||      ||||||||||||   :::
     173 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 222

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     223 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 272

      88 uValAlaCys 91
            :::|||
     273 GCAGACCTGC 282

seq_name: gb_est14:AF063548

seq_documentation_block:
LOCUS        AF063548      511 bp    mRNA            EST       16-JUN-1999
DEFINITION   AF063548 Homo sapiens library (Yu Y) Homo sapiens cDNA clone HA0104
             , mRNA sequence.
ACCESSION    AF063548
VERSION      AF063548.1  GI:5080918
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Yu,Y., Dong,C. and He,F.
  TITLE     Finding some ESTs
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Yongtao Yu
            Department of Hematology
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-68159479
            Fax: 0086-10-68214653
            Email: yyt48@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HA0104"
                     /clone_lib="Homo sapiens library (Yu Y)"
BASE COUNT      114 a    140 c    155 g    102 t
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AF063548   ..

 Align seg 1/1  to: AF063548  from: 1  to: 511

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||             ::
     181 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 215

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     216 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 265

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           |||||||||||||   ||||||      ||||||||||   :::
     266 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 315

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||   ||||||   |||||||||      :::   |||
     316 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 365

      88 uValAlaCys 91
            :::|||
     366 GCAGACCTGC 375

 seq_name: gb_est16:AI110674

 seq_documentation_block:
 LOCUS       AI110674      511 bp    mRNA            EST       11-NOV-1999
 DEFINITION  HA0104 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
             sequence.
 ACCESSION   AI110674
 VERSION     AI110674.1  GI:6359539
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 511)
   AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
             and He,F.
   TITLE     Expression profile analysis of a human fetal liver cDNA library
   JOURNAL   Unpublished (1998)
 COMMENT     Contact: Yongtao Yu
             Department of Hematology
             Beijing Institute of Radiation Medicine
             27 Taiping Road, Beijing 100850, P.R.China
             Tel: 0086-10-68159479
             Fax: 0086-10-68214653
             Email: yyt48@yahoo.com.
 FEATURES             Location/Qualifiers
      source          1. .511
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="Human fetal liver cDNA library"
                      /tissue_type="liver"
                      /dev_stage="fetal"
                      /lab_host="MC1061/P3"
                      /note="Vector: pCDNA1"
 BASE COUNT      114 a    140 c    155 g    102 t
 ORIGIN


 alignment_scores:
             Quality:  129.50             Length:     70
               Ratio:  3.083                Gaps:      2
  Percent Similarity:  60.000   Percent Identity: 41.429

 alignment_block:
  US-09-215-435-177 x AI110674   ..

  Align seg 1/1  to: AI110674  from: 1  to: 511

       24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
          ||||||      |||:::::::::::::::|||             ::
      181 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 215

       38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
          :|||   |||:::|||      :::   |||   |||      :::
      216 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 265

       55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
            |||||||||||||   ||||||      ||||||||||   :::
      266 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 315

       72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
          :::|||||   ||||||   |||||||||      :::   |||
      316 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 365

       88 uValAlaCys 91
             :::|||
      366 GCAGACCTGC 375

  seq_name: gb_est17:AI207711

  seq_documentation_block:
  LOCUS       AI207711      521 bp    mRNA            EST       11-NOV-1999
  DEFINITION  HA3093 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
              sequence.
  ACCESSION   AI207711
  VERSION     AI207711.1  GI:6361726
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 521)
    AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
              and He,F.
    TITLE     Expression profile analysis of a human fetal liver cDNA library
    JOURNAL   Unpublished (1998)
  COMMENT     Contact: Chenggang Zhang
              Beijing Institute of Radiation Medicine
              27 Taiping Road, Beijing 100850, P.R.China
              Email: zhang_chenggang@hotmail.com.
  FEATURES             Location/Qualifiers
       source          1. .521
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
```

```
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      116 a    142 c    157 g    106 t
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AI207711   ..

 Align seg 1/1  to: AI207711  from: 1  to: 521

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||             ::
     181 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 215

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     216 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 265

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           |||||||||||||||   ||||||      ||||||||||   :::
     266 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 315

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||   ||||||   |||||||||      :::   |||
     316 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 365

      88 uValAlaCys 91
            :::|||
     366 GCAGACCTGC 375

seq_name: gb_est17:AI207590

seq_documentation_block:
LOCUS       AI207590      571 bp    mRNA            EST       11-NOV-1999
DEFINITION  HA2954 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI207590
VERSION     AI207590.1  GI:6361600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Chenggang Zhang
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Email: zhang_chenggang@hotmail.com.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      113 a    174 c    149 g    135 t
ORIGIN
```

```
alignment_scores:
            Quality:  129.50             Length:     70
              Ratio:  3.083                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AI207590/rev   ..

 Align seg 1/1  to reverse of: AI207590  from: 1  to: 571

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||             ::
     523 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA..............GA 489

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     488 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 439

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           |||||||||||||||   ||||||      ||||||||||   :::
     438 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 389

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||   ||||||   |||||||||      :::   |||
     388 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 339

      88 uValAlaCys 91
            :::|||
     338 GCAGACCTGC 329

seq_name: gb_est40:AV658823

seq_documentation_block:
LOCUS       AV658823      626 bp    mRNA            EST       07-SEP-2000
DEFINITION  AV658823 GLC Homo sapiens cDNA clone GLCFQE08 3', mRNA sequence.
ACCESSION   AV658823
VERSION     AV658823.1  GI:9879837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCFQE08"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      163 a    157 c    183 g    120 t      3 others
ORIGIN
```

```
alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AV658823   ..

 Align seg 1/1  to: AV658823  from: 1  to: 626

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||              ::
      86 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 120

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     121 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 170

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           |||||||||||||||   ||||||      ||||||||||||   :::
     171 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 220

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||||   ||||||   |||||||||      :::   |||
     221 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 270

      88 uValAlaCys 91
            :::|||
     271 GCAGACCTGC 280

seq_name: gb_est25:AI807643

seq_documentation_block:
LOCUS       AI807643      644 bp    mRNA            EST       19-DEC-1999
DEFINITION  wf49d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2358919 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN
            );, mRNA sequence.
ACCESSION   AI807643
VERSION     AI807643.1  GI:5394209
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1328   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2358919"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      128 a    195 c    173 g    148 t
ORIGIN


alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AI807643/rev   ..

 Align seg 1/1  to reverse of: AI807643  from: 1  to: 644

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||              ::
     520 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 486

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     485 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 436

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           |||||||||||||||   ||||||      ||||||||||||   :::
     435 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 386

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::|||||||   ||||||   |||||||||      :::   |||
     385 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 336

      88 uValAlaCys 91
            :::|||
     335 GCAGACCTGC 326

seq_name: gb_est15:AI065076

seq_documentation_block:
LOCUS       AI065076      662 bp    mRNA            EST       11-NOV-1999
DEFINITION  HA0910 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI065076
VERSION     AI065076.1  GI:6359348
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Yongtao Yu
            Department of Hematology
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-68159479
            Fax: 0086-10-68214653
            Email: yyt48@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      161 a    177 c    182 g    138 t      4 others
ORIGIN
```

```
alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AI065076   ..

 Align seg 1/1  to: AI065076  from: 1  to: 662

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||               ::
     368 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 402

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     403 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 452

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||||   ||||||      ||||||||||||   :::
     453 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 502

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     503 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 552

      88 uValAlaCys 91
            :::|||
     553 GCAGACCTGC 562

seq_name: gb_est40:AV651758

seq_documentation_block:
LOCUS       AV651758      671 bp    mRNA            EST       07-SEP-2000
DEFINITION  AV651758 GLC Homo sapiens cDNA clone GLCCTC02 3', mRNA sequence.
ACCESSION   AV651758
VERSION     AV651758.1  GI:9872772
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCCTC02"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      156 a    183 c    193 g    137 t      2 others
ORIGIN
```

```
alignment_scores:
            Quality:  129.50              Length:     70
              Ratio:  3.083                 Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 41.429

alignment_block:
 US-09-215-435-177 x AV651758   ..

 Align seg 1/1  to: AV651758  from: 1  to: 671

      24 GlyGly......ValAsnLysIleAlaGluLysIleCysGlyAspLeuLy 38
         ||||||      |||:::::::::::::::|||               ::
     241 GGTGGGCAACTGGTAACTGAAGTCACCAAGAAA.............GA 275

      38 sAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisP 55
         :|||   |||:::|||      :::   |||   |||      :::
     276 AGATTCCTGCCAGCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCA 325

      55 heArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPhe 71
           ||||||||||||||||   ||||||      ||||||||||||   :::
     326 GCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTAC 375

      72 SerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGl 88
         :::||||||   ||||||   |||||||||      :::   |||
     376 GGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCT 425

      88 uValAlaCys 91
            :::|||
     426 GCAGACCTGC 435
```



OM of: US-09-215-435-177_COPY_1_24 to: GenEmbl:* out_format : pfs

Date: Nov 15, 2001 12:40 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_pr9:HSDJ601O1	+	120.00	266.48	1.3e-06	33458	! AL109656 Human DNA sequence f
gb_htg9:AC023971	-	120.00	253.75	6.8e-06	153788	! AC023971 Homo sapiens chromc
gb_htg10:AC024934	-	67.00	117.72	255.00	209777	! AC024934 Homo sapiens chromc
gb_htg3:AC012143	-	66.00	117.23	271.32	164426	! AC012143 Homo sapiens clone
gb_pr2:AC008275	+	66.00	117.13	274.99	166504	! AC008275 Homo sapiens BAC cl
gb_htg22:AL451047	-	62.00	107.31	968.73	161545	! AL451047 Homo sapiens chromc
gb_htg13:AC055798	-	62.00	107.02	1.0e+03	167314	! AC055798 Homo sapiens chromc
gb_htg4:AC014616	+	60.00	104.74	1.3e+03	120188	! AC014616 Drosophila melanoga
gb_htg2:AC010577	+	60.00	103.98	1.5e+03	131684	! AC010577 Drosophila melanoga
gb_htg10:AC025616	-	60.00	103.61	1.6e+03	137670	! AC025616 Homo sapiens clone
gb_in2:AE003846	+	60.00	97.15	3.6e+03	298469	! AE003846 Drosophila melanoga
gb_htg7:AC021204	-	59.50	101.18	2.1e+03	158357	! AC021204 Homo sapiens chromc
gb_htg8:AC022211	+	59.00	98.35	3.1e+03	191151	! AC022211 Homo sapiens chromc
gb_pr1:AC005002	-	58.00	99.73	2.6e+03	119841	! AC005002 Homo sapiens clone
gb_pr9:HSJ612B15	-	58.00	99.19	2.7e+03	127820	! AL049597 Human DNA sequence
gb_htg21:AL355981	+	58.00	97.54	3.4e+03	155871	! AL355981 Homo sapiens chromc
gb_htg1:AC004688	+	58.00	96.82	3.7e+03	169794	! AC004688 Plasmodium falcipar
gb_htg1:AC004709	+	58.00	95.62	4.3e+03	196149	! AC004709 Plasmodium falcipar
gb_in3:CEF55A11	-	57.50	110.73	624.36	27598	! Z72511 Caenorhabditis elegans
gb_ro1:AF084363	-	57.00	98.02	3.2e+03	108752	! AF084363 Mus musculus D6MM5e
gb_htg18:AC087882	+	57.00	96.19	4.0e+03	135445	! AC087882 Homo sapiens chromc
gb_htg19:AF282854	+	57.00	96.10	4.1e+03	137020	! AF282854 Homo sapiens chromc
gb_ro1:AC003061	+	57.00	94.66	4.9e+03	162691	! AC003061 Mouse Chromosome 6
gb_pr2:AC007876	-	57.00	94.55	5.0e+03	164879	! AC007876 Homo sapiens BAC cl
gb_htg10:AC024931	+	57.00	93.79	5.5e+03	180519	! AC024931 Homo sapiens chromc
gb_pr3:AC017082	+	57.00	93.64	5.6e+03	183783	! AC017082 Homo sapiens clone
gb_htg5:AC016679	+	57.00	93.47	5.7e+03	187695	! AC016679 Homo sapiens chromc
gb_htg1:AC007305	+	57.00	92.25	6.7e+03	217115	! AC007305 Mus musculus chromc
gb_sts1:CNS070C4	-	56.00	134.49	29.68	1021	! AL423434 T3 end of clone AZ0AA
gb_htg3:AC011770	+	56.00	100.93	2.2e+03	56774	! AC011770 Homo sapiens chromos
gb_htg18:AC090257	-	56.00	99.36	2.7e+03	68585	! AC090257 Homo sapiens chromos
gb_pr1:AC004024	-	56.00	98.16	3.1e+03	79164	! AC004024 Homo sapiens 12q24.2
gb_htg11:AC027448	-	56.00	98.11	3.2e+03	79595	! AC027448 Homo sapiens chromos
gb_in1:AC024854	+	56.00	97.73	3.3e+03	83297	! AC024854 Caenorhabditis elega
gb_in3:CELY71F9B	+	56.00	97.73	3.3e+03	83297	! AC024201 Caenorhabditis elega
gb_htg1:AC008947	+	56.00	97.19	3.5e+03	88948	! AC008947 Homo sapiens chromos
gb_pr3:AC008860	-	56.00	95.73	4.3e+03	105886	! AC008860 Homo sapiens chromc
gb_htg10:AC025444	-	56.00	95.53	4.4e+03	108405	! AC025444 Homo sapiens chromc
gb_htg2:AC010044	-	56.00	94.70	4.9e+03	119844	! AC010044 Drosophila melanoga
gb_pl1:AB001684	-	56.00	92.79	6.2e+03	150613	! AB001684 Chlorella vulgaris
gb_htg23:AL590558	+	56.00	92.59	6.4e+03	154269	! AL590558 Homo sapiens chromc
gb_htg14:AC068206	+	56.00	92.11	6.8e+03	163260	! AC068206 Homo sapiens chr
gb_htg11:AC026360	+	56.00	91.92	7.0e+03	167206	! AC026360 Homo sapiens chr
gb_htg7:AC020687	-	56.00	91.79	7.1e+03	169635	! AC020687 Homo sapiens chr
gb_htg19:AL133513	-	56.00	91.62	7.2e+03	173218	! AL133513 Homo sapiens chr

seq_name: gb_pr9:HSDJ601O1

seq_documentation_block:

```
LOCUS       HSDJ601O1   33458 bp    DNA             PRI       04-APR-2001
DEFINITION  Human DNA sequence from clone RP4-601O1 on chromosome 20. Contains
            the gene for a novel protein with Kunitz/Bovine pancreatic trypsin
            inhibitor domain, the 5' end of a novel gene, part of the gene for
            a novel protein with WAP-type (Whey Acidic Protein) 'four-disulfide
            core' domain, ESTs and GSSs, complete sequence.
ACCESSION   AL109656
VERSION     AL109656.10  GI:10944958
KEYWORDS    HTG; Kunitz; WAP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 33458)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 21, 2000 this sequence version replaced gi:9795206.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-601O1 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-447F3 is at 33359 in this sequence.
            The true right end of clone RP4-688G8 is at 100 in this sequence.
            RP4-601O1 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .33458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP4-601O1"
                     /clone_lib="RPCI-4"
     repeat_region   3. .225
                     /note="L1M4 repeat: matches 3977. .4203 of consensus"
     repeat_region   225. .428
                     /note="L1M4 repeat: matches 4321. .4525 of consensus"
     repeat_region   429. .696
                     /note="AluJb repeat: matches 2. .302 of consensus"
     repeat_region   697. .778
```

```
                     /note="L1M4 repeat: matches 4525. .4606 of consensus"
repeat_region        779. .1089
                     /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region        1090. .1146
                     /note="L1M4 repeat: matches 4606. .4662 of consensus"
repeat_region        1165. .1293
                     /note="LTR42 repeat: matches 1. .134 of consensus"
repeat_region        1351. .1730
                     /note="L1MB5 repeat: matches 5005. .5402 of consensus"
repeat_region        1731. .2032
                     /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region        2056. .2547
                     /note="L1MB5 repeat: matches 5374. .5865 of consensus"
repeat_region        2543. .2729
                     /note="L1MC/D repeat: matches 5428. .5609 of consensus"
repeat_region        2730. .3022
                     /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region        3023. .3181
                     /note="L1MC/D repeat: matches 5609. .5769 of consensus"
repeat_region        3361. .3895
                     /note="L1 repeat: matches 3599. .4167 of consensus"
repeat_region        3986. .4281
                     /note="L2 repeat: matches 534. .834 of consensus"
repeat_region        4526. .4828
                     /note="AluSx repeat: matches 1. .300 of consensus"
misc_feature         5372. .5726
                     /note="match: GSS: Em:AQ076126"
repeat_region        5394. .5470
                     /note="MIR repeat: matches 48. .142 of consensus"
gene                 7401. .7609
                     /gene="dJ601O1.3"
CDS                  <7401. .>7609
                     /gene="dJ601O1.3"
                     /note="possibly a pseudogene
                     match: proteins: Sw:Q14508 Sw:P00993 Sw:Q28631 Sw:P14730
                     Tr:O44131 Wp:C08G9"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="dJ601O1.3 (novel protein with WAP-type (Whey
                     Acidic Protein) 'four-disulfide core' domain)"
                     /protein_id="CAC36024.1"
                     /db_xref="GI:13559237"
                     /translation="FPGSSLGFLSSEHYSSLLVRPFSYLFCVPEYILEPPPCISAPEN
                     CTHLCTMQEDCEKGFQCCSSFCGIV"
mRNA                 complement(7623. .7780)
                     /gene="dJ688G8.4"
                     /note="Continues in Em:AL031671 as dJ688G8.4
                     match: ESTs: Em:AI808149 Em:AI808159"
                     /evidence=not_experimental
                     /product="dJ601O1.2 (a putative novel protein)"
gene                 complement(7623. .7780)
                     /gene="dJ688G8.4"
repeat_region        8352. .8422
                     /note="L2 repeat: matches 2596. .2670 of consensus"
repeat_region        8434. .8656
                     /note="MIR repeat: matches 9. .260 of consensus"
repeat_region        9290. .9419
                     /note="MIR repeat: matches 47. .183 of consensus"
misc_feature         complement(9433. .9899)
                     /note="match: GSS: Em:AQ476416"
repeat_region        9538. .9830
                     /note="AluSg repeat: matches 1. .293 of consensus"
misc_feature         9923. .10651
                     /note="match: GSS: Em:AQ478366"
repeat_region        10453. .10647
                     /note="MER3 repeat: matches 1. .204 of consensus"
misc_feature         complement(10590. .10982)
                     /note="match: GSS: Em:AQ223524"
repeat_region        11007. .11080
                     /note="MER69 repeat: matches 2. .77 of consensus"
repeat_region        11529. .11954
                     /note="MER69 repeat: matches 2070. .2511 of consensus"
repeat_region        13092. .13158
                     /note="L2 repeat: matches 2666. .2732 of consensus"
repeat_region        13202. .13331
                     /note="AluSg/x repeat: matches 183. .312 of consensus"
repeat_region        13332. .15510
                     /note="L1MB1 repeat: matches 3916. .6171 of consensus"
repeat_region        15534. .15832
                     /note="AluSg repeat: matches 1. .299 of consensus"
repeat_region        15846. .16746
                     /note="L1M4 repeat: matches 3472. .4394 of consensus"
repeat_region        16739. .18594
                     /note="L1PA11 repeat: matches 11. .1921 of consensus"
repeat_region        18595. .18888
                     /note="AluSq repeat: matches 1. .295 of consensus"
repeat_region        18889. .20972
                     /note="L1PA11 repeat: matches 1921. .3456 of consensus"
repeat_region        20973. .21280
                     /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region        21281. .23911
                     /note="L1PA11 repeat: matches 3456. .6164 of consensus"
repeat_region        23926. .24129
                     /note="L1M4 repeat: matches 4390. .4613 of consensus"
repeat_region        24130. .24438
                     /note="AluSx repeat: matches 1. .310 of consensus"
repeat_region        24439. .24513
                     /note="L1M4 repeat: matches 4613. .4683 of consensus"
repeat_region        24790. .25008
                     /note="MIR repeat: matches 1. .242 of consensus"
mRNA                 join(25397. .25528,26926. .27103,28675. .28876)
                     /gene="dJ601O1.1"
                     /note="match: cDNAs: Em:AK020315
                     match: ESTs: Em:AW303354"
                     /evidence=not_experimental
                     /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                     pancreatic trypsin inhibitor domain)"
gene                 25397. .28876
                     /gene="dJ601O1.1"
CDS                  join(<25397. .25528,26926. .27103,28675. .28681)
                     /gene="dJ601O1.1"
                     /note="match: proteins: Sw:P08592 Tr:Q13793 Sw:P05067
                     Sw:P12023 Sw:Q95241 Tr:Q9TUI0 Tr:Q91963 Sw:Q02445
                     Tr:O54819 Tr:Q9Z2U8 Sw:P12111 Tr:Q9XWX5 Tr:O61893
                     Sw:P00974 Sw:P04815 Tr:O16701 Sw:P00975 Tr:O70160
                     Sw:P02760 Sw:P48307 Wp:C37C3A Sw:P00991 Sw:P00978
                     Sw:P00976 Sw:P13371 Sw:P04365 Tr:Q9VQT9 Tr:O95925
                     Sw:P00993 Sw:P10280 Tr:Q9TWF8 Tr:Q9TWG0 Tr:Q9TWF9
                     Sw:P26228 Sw:P00980 Sw:P00992 Sw:P00979 Sw:P81658
                     Sw:P25660 Sw:P19761 Sw:P24541 Tr:Q9U8G8 Tr:Q14594
                     Tr:O95428 Tr:O35536"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                     pancreatic trypsin inhibitor domain)"
                     /protein_id="CAC36025.1"
                     /db_xref="GI:13559238"
                     /translation="AACCTMKSAKLGFLLRFFIFCSLNTLLLGGVNKIAEKICGDLKD
                     PCKLDMNFGSCYEVHFRYFYNRTSKRCETFVFSGCNGNLNNFKLKIEREVACVAKYKP
                     PR"
repeat_region        26066. .26228
                     /note="MIR repeat: matches 53. .205 of consensus"
repeat_region        27253. .27579
                     /note="AluSp repeat: matches 1. .303 of consensus"
repeat_region        27649. .27728

alignment_scores:
            Quality:  120.00              Length:      24
              Ratio:  5.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x HSDJ601O1   ..

 Align seg 1/1  to: HSDJ601O1  from: 1  to: 33458
```

```
      1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  25414 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 25463

     17 rLeuAsnThrLeuLeuLeuGly 24
        ||||||||||||||||||||||
  25464 ATTGAATACCCTGTTATTGGGT 25485

seq_name: gb_htg9:AC023971

seq_documentation_block:
LOCUS       AC023971   153788 bp    DNA             HTG       07-APR-2000
DEFINITION  Homo sapiens chromosome 20 clone RP11-770N8 map 20, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC023971
VERSION     AC023971.2  GI:7523758
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 20, clone RP11-770N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 7, 2000 this sequence version replaced gi:7008875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6573
                Center clone name: 770_N_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145038 bases at least Q40
                Consensus quality: 148490 bases at least Q30
                Consensus quality: 150621 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 152588; sum-of-contigs
                Quality coverage: 5.2 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1333: contig of 1333 bp in length
            *    1334 1433: gap of      100 bp
            *    1434     3969: contig of 2536 bp in length
            *    3970 4069: gap of      100 bp
            *    4070     7211: contig of 3142 bp in length
            *    7212 7311: gap of      100 bp
            *    7312    13518: contig of 6207 bp in length
            *   13519 13618: gap of      100 bp
            *   13619    20876: contig of 7258 bp in length
            *   20877 20976: gap of      100 bp
            *   20977    28838: contig of 7862 bp in length
            *   28839 28938: gap of      100 bp
            *   28939    37590: contig of 8652 bp in length
            *   37591 37690: gap of      100 bp
            *   37691    45814: contig of 8124 bp in length
            *   45815 45914: gap of      100 bp
            *   45915    62601: contig of 16687 bp in length
            *   62602 62701: gap of      100 bp
            *   62702    81817: contig of 19116 bp in length
            *   81818 81917: gap of      100 bp
            *   81918   100273: contig of 18356 bp in length
            *  100274 100373: gap of      100 bp
            *  100374   125266: contig of 24893 bp in length
            *  125267 125366: gap of      100 bp
            *  125367   153788: contig of 28422 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20"
                     /clone="RP11-770N8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .3969
                     /note="assembly_fragment"
     misc_feature    4070. .7211
                     /note="assembly_fragment"
     misc_feature    7312. .13518
                     /note="assembly_fragment"
     misc_feature    13619. .20876
                     /note="assembly_fragment"
     misc_feature    20977. .28838
                     /note="assembly_fragment"
     misc_feature    28939. .37590
                     /note="assembly_fragment"
     misc_feature    37691. .45814
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    45915. .62601
                     /note="assembly_fragment"
     misc_feature    62702. .81817
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    81918. .100273
                     /note="assembly_fragment"
     misc_feature    100374. .125266
                     /note="assembly_fragment"
     misc_feature    125367. .153788
                     /note="assembly_fragment"
BASE COUNT    44583 a  30904 c  30568 g  46532 t   1201 others
```

```
ORIGIN


alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC023971/rev   ..

 Align seg 1/1  to reverse of: AC023971  from: 1  to: 153788

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  135635 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 135586

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
  135585 ATTGAATACCCTGTTATTGGGT 135564

seq_name: gb_htg10:AC024934

seq_documentation_block:
LOCUS       AC024934   209777 bp    DNA             HTG       30-DEC-2000
DEFINITION  Homo sapiens chromosome 12 clone RP11-329A19, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC024934
VERSION     AC024934.29  GI:11995487
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209777)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209777)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 29, 2000 this sequence version replaced gi:11192044.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: HAJA
                Center clone name: RP11-329A19
            ------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 8% of reads
                Chemistry: Dye-terminator Big Dye: 92% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 197273 bases at least Q40
                Consensus quality: 202546 bases at least Q30
                Consensus quality: 205790 bases at least Q20
                Estimated insert size: 202671; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 5.5x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33513: contig of 33513 bp in length
            *    33514    33613: gap of unknown length
            *    33614    58205: contig of 24592 bp in length
            *    58206    58305: gap of unknown length
            *    58306    91615: contig of 33310 bp in length
            *    91616    91715: gap of unknown length
            *    91716   114617: contig of 22902 bp in length
            *   114618   114717: gap of unknown length
            *   114718   132860: contig of 18143 bp in length
            *   132861   132960: gap of unknown length
            *   132961   152239: contig of 19279 bp in length
            *   152240   152339: gap of unknown length
            *   152340   169844: contig of 17505 bp in length
            *   169845   169944: gap of unknown length
            *   169945   185251: contig of 15307 bp in length
            *   185252   185351: gap of unknown length
            *   185352   194558: contig of 9207 bp in length
            *   194559   194658: gap of unknown length
            *   194659   204080: contig of 9422 bp in length
            *   204081   204180: gap of unknown length
            *   204181   206129: contig of 1949 bp in length
            *   206130   206229: gap of unknown length
            *   206230   207330: contig of 1101 bp in length.
            *   207331   207430: gap of unknown length
            *   207431   208674: contig of 1244 bp in length
            *   208675   208774: gap of unknown length
            *   208775   209777: contig of 1003 bp in length.
FEATURES             Location/Qualifiers
     source          1. .209777
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-329A19"
```

```
BASE COUNT    65720 a  39190 c  38969 g  64529 t   1369 others
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     22
              Ratio:  4.188               Gaps:      1
 Percent Similarity:  72.727   Percent Identity: 68.182

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC024934/rev   ..

 Align seg 1/1  to reverse of: AC024934  from: 1  to: 209777

       7 GlyPheLeuLeuArgPhePhe.....IlePheCysSerLeuAs 19
         ||||||||||||||||||||           ||||||||||||||||
   45250 GGTTTTCTTTTAAGATTTTTTTCCTTATCTTTGATTTTCTGCAGTTTGAA 45201

      19 nThrLeuLeuLeuGly 24
         ||||:::      |||
   45200 TACAATTTGCCCAGGT 45185

seq_name: gb_htg3:AC012143

seq_documentation_block:
LOCUS       AC012143   164426 bp    DNA             HTG       10-SEP-2000
DEFINITION  Homo sapiens clone RP11-11G20, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC012143
VERSION     AC012143.3  GI:10047745
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164426)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-11G20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164426)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:7230063.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3159
                Center clone name: 11_G_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158797 bases at least Q40
                Consensus quality: 161636 bases at least Q30
                Consensus quality: 162651 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 163726; sum-of-contigs
                Quality coverage: 5.7 in Q20 bases; agarose-fp
                Quality coverage: 5.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    33254: contig of 33254 bp in length
            *    33255 33354: gap of      100 bp
            *    33355    37499: contig of 4145 bp in length
            *    37500 37599: gap of      100 bp
            *    37600    40875: contig of 3276 bp in length
            *    40876 40975: gap of      100 bp
            *    40976    91250: contig of 50275 bp in length
            *    91251 91350: gap of      100 bp
            *    91351   107305: contig of 15955 bp in length
            *   107306 107405: gap of      100 bp
            *   107406   124992: contig of 17587 bp in length
            *   124993 125092: gap of      100 bp
            *   125093   147928: contig of 22836 bp in length
            *   147929 148028: gap of      100 bp
            *   148029   164426: contig of 16398 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-11G20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .33254
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    33355. .37499
                     /note="assembly_fragment"
     misc_feature    37600. .40875
                     /note="assembly_fragment"
     misc_feature    40976. .91250
                     /note="assembly_fragment"
     misc_feature    91351. .107305
                     /note="assembly_fragment"
     misc_feature    107406. .124992
                     /note="assembly_fragment"
     misc_feature    125093. .147928
                     /note="assembly_fragment"
     misc_feature    148029. .164426
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    52060 a  30107 c  29871 g  51686 t    702 others
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     23
              Ratio:  3.143               Gaps:      0
 Percent Similarity:  91.304   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC012143/rev   ..

 Align seg 1/1  to reverse of: AC012143  from: 1  to: 164426

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         |||:::      |||:::::::|||||||||:::::::::::|||||||
```

```
   24972 ATGCAAGTAATCAAATTTTCCTTCCTGTTAAAGATGCTCTTATTTTGTTC 24923

      17 rLeuAsnThrLeuLeuLeu 23
         |||||||:::|||:::|||
   24922 TCTAAATTCTTTGATTCTT 24904

seq_name: gb_pr2:AC008275

seq_documentation_block:
LOCUS       AC008275   166504 bp    DNA             PRI       30-SEP-2000
DEFINITION  Homo sapiens BAC clone RP11-294I11 from 2, complete sequence.
ACCESSION   AC008275
VERSION     AC008275.3  GI:6598922
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166504)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 166504)
  AUTHORS   Hou,S., Stoneking,T., Yoakum,M. and Tucci,S.
  TITLE     The sequence of Homo sapiens BAC clone RP11-294I11
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 166504)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 166504)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 166504)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:5757528.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0294I11
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-465G21.  Actual start of
            this clone is at base position 1 of RP11-294I11; actual end is at
            base position 166504 of RP11-294I11.
FEATURES                Location/Qualifiers
     source             1. .166504
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-294I11"
                        /clone_lib="RPCI-11"
     repeat_region      1051. .1346
                        /rpt_family="Alu"
     repeat_region      1375. .1538
                        /rpt_family="L1"
     repeat_region      2277. .2522
                        /rpt_family="Alu"
     repeat_region      2523. .2580
                        /rpt_family="(TAAA)n"
     repeat_region      2603. .3434
                        /rpt_family="L1"
     repeat_region      3504. .3955
                        /rpt_family="L1"
     repeat_region      3956. .4238
                        /rpt_family="Alu"
     repeat_region      4239. .4311
                        /rpt_family="L1"
     repeat_region      4514. .4595
                        /rpt_family="(TATG)n"
     repeat_region      4736. .4759
                        /rpt_family="AT_rich"
     repeat_region      5357. .5645
                        /rpt_family="MER4-group"
     repeat_region      5647. .5941
                        /rpt_family="L1"
     repeat_region      5977. .5999
                        /rpt_family="(TA)n"
     repeat_region      6012. .6289
                        /rpt_family="MER4-group"
     repeat_region      6793. .12276
                        /rpt_family="L1"
     repeat_region      15813. .16189
                        /rpt_family="MaLR"
     repeat_region      16572. .16605
                        /rpt_family="(CA)n"
     repeat_region      17466. .17829
                        /rpt_family="MaLR"
     repeat_region      18945. .18979
                        /rpt_family="A-rich"
     repeat_region      19114. .19222
                        /rpt_family="MIR"
     repeat_region      19368. .19475
                        /rpt_family="MIR"
     repeat_region      19479. .19518
                        /rpt_family="AT_rich"
     repeat_region      20420. .20475
                        /rpt_family="AT_rich"
     repeat_region      22046. .22169
                        /rpt_family="MIR"
     repeat_region      22214. .22246
                        /rpt_family="(GAAAA)n"
     repeat_region      22722. .23007
```

```
                /rpt_family="Alu"
     repeat_region   24143. .24649
                /rpt_family="L2"
     repeat_region   24650. .24945
                /rpt_family="Alu"
     repeat_region   24946. .25143
                /rpt_family="L2"
     repeat_region   25470. .25780
                /rpt_family="Alu"
     repeat_region   26007. .26209
                /rpt_family="L2"
     repeat_region   26210. .26330
                /rpt_family="Alu"
     repeat_region   26346. .26919
                /rpt_family="L2"
     repeat_region   27078. .27249
                /rpt_family="MaLR"
     repeat_region   27398. .27446
                /rpt_family="L2"
     repeat_region   27711. .27752
                /rpt_family="GA-rich"
     repeat_region   28051. .28242
                /rpt_family="MIR"
     repeat_region   28273. .28378
                /rpt_family="MIR"
     repeat_region   28415. .28782
                /rpt_family="MaLR"
     repeat_region   28783. .29074
                /rpt_family="MaLR"
     repeat_region   29078. .29186
                /rpt_family="MaLR"
     repeat_region   30033. .30125
                /rpt_family="(TATG)n"
     repeat_region   30150. .30635
                /rpt_family="MaLR"
     repeat_region   31305. .31436
                /rpt_family="L2"
     repeat_region   31814. .32012
                /rpt_family="MaLR"
     repeat_region   32053. .32073
                /rpt_family="AT_rich"
     repeat_region   33095. .34300
                /rpt_family="L1"
     repeat_region   34293. .34927
                /rpt_family="L1"
     repeat_region   34928. .35684
                /rpt_family="Retroviral"
     repeat_region   35698. .35848
                /rpt_family="L1"
     repeat_region   36036. .36312
                /rpt_family="Alu"
     repeat_region   36313. .36343
                /rpt_family="(CAAA)n"
     repeat_region   36344. .36365
                /rpt_family="Alu"
     repeat_region   37140. .37856
                /rpt_family="Other"
     repeat_region   38227. .38357
                /rpt_family="Alu"
     repeat_region   38358. .38641
                /rpt_family="Alu"
     repeat_region   38733. .38759
                /rpt_family="AT_rich"

alignment_scores:
           Quality:   66.00            Length:     23
             Ratio:  3.143              Gaps:      0
 Percent Similarity:  91.304   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC008275   ..

 Align seg 1/1  to: AC008275  from: 1  to: 166504
```

```
       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         |||:::      |||:::::::|||||||||:::::::::::::|||||||
   51082 ATGCAAGTAATCAAATTTTCCTTCCTGTTAAAGATGCTCTTATTTTGTTC 51131

      17 rLeuAsnThrLeuLeuLeu 23
         |||||||:::|||:::|||
   51132 TCTAAATTCTTTGATTCTT 51150

seq_name: gb_htg22:AL451047

seq_documentation_block:
LOCUS       AL451047   161545 bp    DNA             HTG       07-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-1B20, *** SEQUENCING IN
            PROGRESS ***, 28 unordered pieces.
ACCESSION   AL451047 AC055798
VERSION     AL451047.1  GI:11602617
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161545)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA1B20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 146641 bases at least Q40
            Consensus quality: 152102 bases at least Q30
            Consensus quality: 155353 bases at least Q20
            Insert size: 158845; sum-of-contigs
            Insert size: 171731; agarose-fp
            Quality coverage: 3.30x in Q20 bases; sum-of-contigs Quality
            coverage: 3.29x in Q20 bases; agarose-fp
            --------------
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4795: contig of 4795 bp in length
            *    4796 4895: gap of      100 bp
            *    4896     9059: contig of 4164 bp in length
            *    9060 9159: gap of      100 bp
            *    9160    12528: contig of 3369 bp in length
            *   12529 12628: gap of      100 bp
            *   12629    26952: contig of 14324 bp in length
            *   26953 27052: gap of      100 bp
            *   27053    31767: contig of 4715 bp in length
            *   31768 31867: gap of      100 bp
            *   31868    36136: contig of 4269 bp in length
            *   36137 36236: gap of      100 bp
            *   36237    39538: contig of 3302 bp in length
            *   39539 39638: gap of      100 bp
```

```
            *    39639    50902: contig of 11264 bp in length
            *    50903 51002: gap of      100 bp
            *    51003    57533: contig of 6531 bp in length
            *    57534 57633: gap of      100 bp
            *    57634    79585: contig of 21952 bp in length
            *    79586 79685: gap of      100 bp
            *    79686    85812: contig of 6127 bp in length
            *    85813 85912: gap of      100 bp
            *    85913    88090: contig of 2178 bp in length
            *    88091 88190: gap of      100 bp
            *    88191    99185: contig of 10995 bp in length
            *    99186 99285: gap of      100 bp
            *    99286   102918: contig of 3633 bp in length
            *   102919 103018: gap of      100 bp
            *   103019   105939: contig of 2921 bp in length
            *   105940 106039: gap of      100 bp
            *   106040   110442: contig of 4403 bp in length
            *   110443 110542: gap of      100 bp
            *   110543   114097: contig of 3555 bp in length
            *   114098 114197: gap of      100 bp
            *   114198   116474: contig of 2277 bp in length
            *   116475 116574: gap of      100 bp
            *   116575   123146: contig of 6572 bp in length
            *   123147 123246: gap of      100 bp
            *   123247   127668: contig of 4422 bp in length
            *   127669 127768: gap of      100 bp
            *   127769   130242: contig of 2474 bp in length
            *   130243 130342: gap of      100 bp
            *   130343   132879: contig of 2537 bp in length
            *   132880 132979: gap of      100 bp
            *   132980   136877: contig of 3898 bp in length
            *   136878 136977: gap of      100 bp
            *   136978   139814: contig of 2837 bp in length
            *   139815 139914: gap of      100 bp
            *   139915   147002: contig of 7088 bp in length
            *   147003 147102: gap of      100 bp
            *   147103   150773: contig of 3671 bp in length
            *   150774 150873: gap of      100 bp
            *   150874   158190: contig of 7317 bp in length
            *   158191 158290: gap of      100 bp
            *   158291   161545: contig of 3255 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-1B20"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .4795
                     /note="assembly_fragment:00003.0"
     misc_feature    4896. .9059
                     /note="assembly_fragment:00051"
     misc_feature    9160. .12528
                     /note="assembly_fragment:00082"
     misc_feature    12629. .26952
                     /note="assembly_fragment:00138"
     misc_feature    27053. .31767
                     /note="assembly_fragment:00162"
     misc_feature    31868. .36136
                     /note="assembly_fragment:00282"
     misc_feature    36237. .39538
                     /note="assembly_fragment:00329"
     misc_feature    39639. .50902
                     /note="assembly_fragment:00331"
     misc_feature    51003. .57533
                     /note="assembly_fragment:00381"
     misc_feature    57634. .79585
                     /note="assembly_fragment:00394"
     misc_feature    79686. .85812
                     /note="assembly_fragment:00423"
     misc_feature    85913. .88090
                     /note="assembly_fragment:00424"
     misc_feature    88191. .99185
                     /note="assembly_fragment:00549"
     misc_feature    99286. .102918
                     /note="assembly_fragment:00592"
     misc_feature    103019. .105939
                     /note="assembly_fragment:00595"
     misc_feature    106040. .110442
                     /note="assembly_fragment:00612"
     misc_feature    110543. .114097
                     /note="assembly_fragment:00640"
     misc_feature    114198. .116474
                     /note="assembly_fragment:00736"
     misc_feature    116575. .123146
                     /note="assembly_fragment:00809"
     misc_feature    123247. .127668
                     /note="assembly_fragment:00834"
     misc_feature    127769. .130242
                     /note="assembly_fragment:00845"
     misc_feature    130343. .132879
                     /note="assembly_fragment:00859"
     misc_feature    132980. .136877
                     /note="assembly_fragment:00877"
     misc_feature    136978. .139814
                     /note="assembly_fragment:00900"
     misc_feature    139915. .147002
                     /note="assembly_fragment:00916"
     misc_feature    147103. .150773
                     /note="assembly_fragment:00954"
     misc_feature    150874. .158190
                     /note="assembly_fragment:01048"
     misc_feature    158291. .161545
                     /note="assembly_fragment:01172"
BASE COUNT    51501 a  29748 c  28475 g  49116 t   2705 others
ORIGIN


alignment_scores:
            Quality:   62.00              Length:     18
              Ratio:  4.133                Gaps:      0
 Percent Similarity:  83.333   Percent Identity: 66.667

alignment_block:
 US-09-215-435-177_COPY_1_24 x AL451047/rev   ..

 Align seg 1/1  to reverse of: AL451047  from: 1  to: 161545

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         |||||||:::|||   |||||||||||||||||||||||:::   :::   ||
    9646 GGTTTTTTTCTTTATTTTTTTATATTCTGTAGTTTGAAAATGGTGTGCCT 9597

      23 uGly 24
         ||||
    9596 AGGT 9593

 seq_name: gb_htg13:AC055798

 seq_documentation_block:
 LOCUS       AC055798   167314 bp    DNA             HTG       15-JUN-2000
 DEFINITION  Homo sapiens chromosome 1 clone RP11-1B20 map 1, WORKING DRAFT
             SEQUENCE, 42 unordered pieces.
 ACCESSION   AC055798
 VERSION     AC055798.3  GI:8567859
 KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 167314)
   AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
   TITLE     Homo sapiens chromosome 1, clone RP11-1B20
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 167314)
   AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
```

```
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 15, 2000 this sequence version replaced gi:8113488.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7563
                Center clone name: 1_B_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 146335 bases at least Q40
                Consensus quality: 156294 bases at least Q30
                Consensus quality: 160582 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 163214; sum-of-contigs
                Quality coverage: 3.2 in Q20 bases; agarose-fp
                Quality coverage: 3.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 42 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1267: contig of 1267 bp in length
            *     1268 1367: gap of      100 bp
            *     1368     2781: contig of 1414 bp in length
            *     2782 2881: gap of      100 bp
            *     2882     3957: contig of 1076 bp in length
            *     3958 4057: gap of      100 bp
            *     4058     5450: contig of 1393 bp in length
            *     5451 5550: gap of      100 bp
            *     5551     6932: contig of 1382 bp in length
            *     6933 7032: gap of      100 bp
            *     7033     8658: contig of 1626 bp in length
            *     8659 8758: gap of      100 bp
            *     8759    10155: contig of 1397 bp in length
            *    10156 10255: gap of      100 bp
            *    10256    11921: contig of 1666 bp in length
            *    11922 12021: gap of      100 bp
            *    12022    13078: contig of 1057 bp in length
            *    13079 13178: gap of      100 bp
            *    13179    14458: contig of 1280 bp in length
            *    14459 14558: gap of      100 bp
            *    14559    15639: contig of 1081 bp in length
            *    15640 15739: gap of      100 bp
            *    15740    18149: contig of 2410 bp in length
            *    18150 18249: gap of      100 bp
            *    18250    21248: contig of 2999 bp in length
            *    21249 21348: gap of      100 bp
            *    21349    23819: contig of 2471 bp in length
            *    23820 23919: gap of      100 bp
            *    23920    27086: contig of 3167 bp in length
            *    27087 27186: gap of      100 bp
            *    27187    29461: contig of 2275 bp in length
            *    29462 29561: gap of      100 bp
            *    29562    32385: contig of 2824 bp in length
            *    32386 32485: gap of      100 bp
            *    32486    35779: contig of 3294 bp in length
            *    35780 35879: gap of      100 bp
            *    35880    39099: contig of 3220 bp in length
            *    39100 39199: gap of      100 bp
            *    39200    42626: contig of 3427 bp in length
            *    42627 42726: gap of      100 bp
            *    42727    46753: contig of 4027 bp in length
            *    46754 46853: gap of      100 bp
            *    46854    50733: contig of 3880 bp in length
            *    50734 50833: gap of      100 bp
            *    50834    54751: contig of 3918 bp in length
            *    54752 54851: gap of      100 bp
            *    54852    57785: contig of 2934 bp in length
            *    57786 57885: gap of      100 bp
            *    57886    60899: contig of 3014 bp in length
            *    60900 60999: gap of      100 bp
            *    61000    64635: contig of 3636 bp in length
            *    64636 64735: gap of      100 bp
            *    64736    68502: contig of 3767 bp in length
            *    68503 68602: gap of      100 bp
            *    68603    71649: contig of 3047 bp in length
            *    71650 71749: gap of      100 bp
            *    71750    76193: contig of 4444 bp in length
            *    76194 76293: gap of      100 bp
            *    76294    80893: contig of 4600 bp in length
            *    80894 80993: gap of      100 bp
            *    80994    87486: contig of 6493 bp in length
            *    87487 87586: gap of      100 bp
            *    87587    91852: contig of 4266 bp in length
            *    91853 91952: gap of      100 bp
            *    91953    97812: contig of 5860 bp in length
            *    97813 97912: gap of      100 bp
            *    97913   102283: contig of 4371 bp in length
            *   102284 102383: gap of      100 bp
            *   102384   108866: contig of 6483 bp in length
            *   108867 108966: gap of      100 bp
            *   108967   114480: contig of 5514 bp in length
            *   114481 114580: gap of      100 bp
            *   114581   121779: contig of 7199 bp in length
            *   121780 121879: gap of      100 bp
            *   121880   126681: contig of 4802 bp in length
            *   126682 126781: gap of      100 bp
            *   126782   133824: contig of 7043 bp in length
            *   133825 133924: gap of      100 bp
            *   133925   144194: contig of 10270 bp in length
            *   144195 144294: gap of      100 bp
            *   144295   155524: contig of 11230 bp in length
            *   155525 155624: gap of      100 bp
            *   155625   167314: contig of 11690 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167314
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-1B20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1267
```

```
                     /note="assembly_fragment"
     misc_feature    1368. .2781
                     /note="assembly_fragment"
     misc_feature    2882. .3957
                     /note="assembly_fragment"
     misc_feature    4058. .5450
                     /note="assembly_fragment"
     misc_feature    5551. .6932
                     /note="assembly_fragment"
     misc_feature    7033. .8658
                     /note="assembly_fragment"
     misc_feature    8759. .10155
                     /note="assembly_fragment"
     misc_feature    10256. .11921
                     /note="assembly_fragment"
     misc_feature    12022. .13078
                     /note="assembly_fragment"
     misc_feature    13179. .14458
                     /note="assembly_fragment"
     misc_feature    14559. .15639
                     /note="assembly_fragment"
     misc_feature    15740. .18149
                     /note="assembly_fragment"
     misc_feature    18250. .21248
                     /note="assembly_fragment"
     misc_feature    21349. .23819
                     /note="assembly_fragment"
     misc_feature    23920. .27086
                     /note="assembly_fragment"
     misc_feature    27187. .29461
                     /note="assembly_fragment"
     misc_feature    29562. .32385
                     /note="assembly_fragment"
     misc_feature    32486. .35779
                     /note="assembly_fragment"
     misc_feature    35880. .39099
                     /note="assembly_fragment"

alignment_scores:
            Quality:   62.00            Length:     18
              Ratio:  4.133              Gaps:      0
 Percent Similarity:  83.333   Percent Identity: 66.667

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC055798/rev   ..

 Align seg 1/1  to reverse of: AC055798  from: 1  to: 167314

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         ||||||:::|||   ||||||||||||||||||||||:::   :::   ||
   24205 GGTTTTTTTCTTTATTTTTTTATATTCTGTAGTTTGAAAATGGTGTGCCT 24156

      23 uGly 24
         ||||
   24155 AGGT 24152

seq_name: gb_htg4:AC014616

seq_documentation_block:
LOCUS       AC014616   120188 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014616
VERSION     AC014616.1  GI:6436719
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 120188)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212336 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .120188
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    39683 a   20457 c   20389 g   39659 t
ORIGIN


alignment_scores:
            Quality:   60.00            Length:     20
              Ratio:  3.750              Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 55.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC014616   ..

 Align seg 1/1  to: AC014616  from: 1  to: 120188

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
   54820 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 54869

      18 uAsnThrLeu 21
         ||||:::|||
   54870 AAATAGCCTA 54879

seq_name: gb_htg2:AC010577

seq_documentation_block:
LOCUS       AC010577   131684 bp    DNA             HTG       13-OCT-1999
DEFINITION  Drosophila melanogaster chromosome 4 clone BACR22J20 (D1076)
            RPCI-98 22.J.20 map 101F-102F strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 58 unordered pieces.
ACCESSION   AC010577
VERSION     AC010577.3  GI:6041726
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 131684)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 131684)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
```

```
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 15, 1999 this sequence version replaced gi:5922055.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 58 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      539: contig of 539 bp in length
            *      540      619: gap of unknown length
            *      620     1252: contig of 633 bp in length
            *     1253     1332: gap of unknown length
            *     1333     2031: contig of 699 bp in length
            *     2032     2111: gap of unknown length
            *     2112     2503: contig of 392 bp in length
            *     2504     2583: gap of unknown length
            *     2584     3666: contig of 1083 bp in length
            *     3667     3746: gap of unknown length
            *     3747     4422: contig of 676 bp in length
            *     4423     4502: gap of unknown length
            *     4503     5556: contig of 1054 bp in length
            *     5557     5636: gap of unknown length
            *     5637     6287: contig of 651 bp in length
            *     6288     6367: gap of unknown length
            *     6368     7159: contig of 792 bp in length
            *     7160     7239: gap of unknown length
            *     7240     8535: contig of 1296 bp in length
            *     8536     8615: gap of unknown length
            *     8616     9729: contig of 1114 bp in length
            *     9730     9809: gap of unknown length
            *     9810    10558: contig of 749 bp in length
            *    10559    10638: gap of unknown length
            *    10639    11610: contig of 972 bp in length
            *    11611    11690: gap of unknown length
            *    11691    12426: contig of 736 bp in length
            *    12427    12506: gap of unknown length
            *    12507    13606: contig of 1100 bp in length
            *    13607    13686: gap of unknown length
            *    13687    15581: contig of 1895 bp in length
            *    15582    15661: gap of unknown length
            *    15662    17324: contig of 1663 bp in length
            *    17325    17404: gap of unknown length
            *    17405    18936: contig of 1532 bp in length
            *    18937    19016: gap of unknown length
            *    19017    20108: contig of 1092 bp in length
            *    20109    20188: gap of unknown length
            *    20189    22142: contig of 1954 bp in length
            *    22143    22222: gap of unknown length
            *    22223    23815: contig of 1593 bp in length
            *    23816    23895: gap of unknown length
            *    23896    25279: contig of 1384 bp in length
            *    25280    25359: gap of unknown length
            *    25360    27692: contig of 2333 bp in length
            *    27693    27772: gap of unknown length
            *    27773    29390: contig of 1618 bp in length
            *    29391    29470: gap of unknown length
            *    29471    31101: contig of 1631 bp in length
            *    31102    31181: gap of unknown length
            *    31182    33910: contig of 2729 bp in length
            *    33911    33990: gap of unknown length
            *    33991    35890: contig of 1900 bp in length
            *    35891    35970: gap of unknown length
            *    35971    38552: contig of 2582 bp in length
            *    38553    38632: gap of unknown length
            *    38633    39807: contig of 1175 bp in length
            *    39808    39887: gap of unknown length
            *    39888    41904: contig of 2017 bp in length
            *    41905    41984: gap of unknown length
            *    41985    45618: contig of 3634 bp in length
            *    45619    45698: gap of unknown length
            *    45699    47674: contig of 1976 bp in length
            *    47675    47754: gap of unknown length
            *    47755    49734: contig of 1980 bp in length
            *    49735    49814: gap of unknown length
            *    49815    53262: contig of 3448 bp in length
            *    53263    53342: gap of unknown length
            *    53343    56427: contig of 3085 bp in length
            *    56428    56507: gap of unknown length
            *    56508    59066: contig of 2559 bp in length
            *    59067    59146: gap of unknown length
            *    59147    61633: contig of 2487 bp in length
            *    61634    61713: gap of unknown length
            *    61714    64993: contig of 3280 bp in length
            *    64994    65073: gap of unknown length
            *    65074    67655: contig of 2582 bp in length
            *    67656    67735: gap of unknown length
            *    67736    70362: contig of 2627 bp in length
            *    70363    70442: gap of unknown length
            *    70443    74287: contig of 3845 bp in length
            *    74288    74367: gap of unknown length
            *    74368    80483: contig of 6116 bp in length
            *    80484    80563: gap of unknown length
            *    80564    86400: contig of 5837 bp in length
            *    86401    86480: gap of unknown length
            *    86481    95655: contig of 9175 bp in length
            *    95656    95735: gap of unknown length
            *    95736   104481: contig of 8746 bp in length
            *   104482   104561: gap of unknown length
            *   104562   114326: contig of 9765 bp in length
            *   114327   114406: gap of unknown length
            *   114407   124091: contig of 9685 bp in length
            *   124092   124171: gap of unknown length
            *   124172   124731: contig of 560 bp in length
            *   124732   124811: gap of unknown length
            *   124812   125551: contig of 740 bp in length
            *   125552   125631: gap of unknown length
            *   125632   126207: contig of 576 bp in length
            *   126208   126287: gap of unknown length
            *   126288   126904: contig of 617 bp in length
            *   126905   126984: gap of unknown length
            *   126985   127691: contig of 707 bp in length
            *   127692   127771: gap of unknown length
            *   127772   128367: contig of 596 bp in length
            *   128368   128447: gap of unknown length
            *   128448   129055: contig of 608 bp in length
            *   129056   129135: gap of unknown length
            *   129136   129672: contig of 537 bp in length
            *   129673   129752: gap of unknown length
            *   129753   130352: contig of 600 bp in length
            *   130353   130432: gap of unknown length
            *   130433   131056: contig of 624 bp in length
            *   131057   131136: gap of unknown length
            *   131137   131684: contig of 548 bp in length.
FEATURES             Location/Qualifiers
     source          1. .131684
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="4"
                     /map="101F-102F"
                     /clone="BACR22J20 (D1076) RPCI-98 22.J.20"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"
BASE COUNT    41744 a  21575 c  22174 g  41614 t   4577 others
ORIGIN


alignment_scores:
```

```
          Quality:   60.00          Length:     20
            Ratio:   3.750            Gaps:      0
Percent Similarity:  80.000  Percent Identity: 55.000

alignment_block:
US-09-215-435-177_COPY_1_24 x AC010577   ..

Align seg 1/1  to: AC010577  from: 1  to: 131684

      2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
        |||   ::::::|||||||   |||||||   |||:::|||:::|||   ||
  41209 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 41258

     18 uAsnThrLeu 21
        ||||:::|||
  41259 AAATAGCCTA 41268

seq_name: gb_htg10:AC025616

seq_documentation_block:
LOCUS       AC025616   137670 bp    DNA              HTG       20-APR-2000
DEFINITION  Homo sapiens clone RP11-21J10, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC025616
VERSION     AC025616.3  GI:7596801
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137670)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21J10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137670)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Apr 19, 2000 this sequence version replaced gi:7331406.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4030
                Center clone name: 21_J_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 128269 bases at least Q40
                Consensus quality: 133012 bases at least Q30
                Consensus quality: 134872 bases at least Q20
                Insert size: 136000; agarose-fp
                Insert size: 136170; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1388: contig of 1388 bp in length
            *    1389  1488: gap of      100 bp
            *    1489     3431: contig of 1943 bp in length
            *    3432  3531: gap of      100 bp
            *    3532     6098: contig of 2567 bp in length
            *    6099  6198: gap of      100 bp
            *    6199     9757: contig of 3559 bp in length
            *    9758  9857: gap of      100 bp
            *    9858    12826: contig of 2969 bp in length
            *   12827 12926: gap of      100 bp
            *   12927    16317: contig of 3391 bp in length
            *   16318 16417: gap of      100 bp
            *   16418    20454: contig of 4037 bp in length
            *   20455 20554: gap of      100 bp
            *   20555    24651: contig of 4097 bp in length
            *   24652 24751: gap of      100 bp
            *   24752    31538: contig of 6787 bp in length
            *   31539 31638: gap of      100 bp
            *   31639    39011: contig of 7373 bp in length
            *   39012 39111: gap of      100 bp
            *   39112    47730: contig of 8619 bp in length
            *   47731 47830: gap of      100 bp
            *   47831    61833: contig of 14003 bp in length
            *   61834 61933: gap of      100 bp
            *   61934    75049: contig of 13116 bp in length
            *   75050 75149: gap of      100 bp
            *   75150    91402: contig of 16253 bp in length
            *   91403 91502: gap of      100 bp
            *   91503   108119: contig of 16617 bp in length
            *  108120 108219: gap of      100 bp
            *  108220   137670: contig of 29451 bp in length.
FEATURES             Location/Qualifiers
     source          1. .137670
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21J10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1388
                     /note="assembly_fragment"
     misc_feature    1489. .3431
                     /note="assembly_fragment"
     misc_feature    3532. .6098
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    6199. .9757
                     /note="assembly_fragment"
     misc_feature    9858. .12826
                     /note="assembly_fragment"
     misc_feature    12927. .16317
                     /note="assembly_fragment"
     misc_feature    16418. .20454
                     /note="assembly_fragment"
     misc_feature    20555. .24651
                     /note="assembly_fragment
```

```
                     clone_end:T7
                     vector_side:right"
     misc_feature    24752. .31538
                     /note="assembly_fragment"
     misc_feature    31639. .39011
                     /note="assembly_fragment"
     misc_feature    39112. .47730
                     /note="assembly_fragment"
     misc_feature    47831. .61833
                     /note="assembly_fragment"
     misc_feature    61934. .75049
                     /note="assembly_fragment"
     misc_feature    75150. .91402
                     /note="assembly_fragment"
     misc_feature    91503. .108119
                     /note="assembly_fragment"
     misc_feature    108220. .137670
                     /note="assembly_fragment"
BASE COUNT    44677 a  24460 c  24057 g  42971 t   1505 others
ORIGIN


alignment_scores:
            Quality:   60.00             Length:     23
              Ratio:  3.158               Gaps:      0
 Percent Similarity:  82.609   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC025616/rev   ..

 Align seg 1/1  to reverse of: AC025616  from: 1  to: 137670

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         |||:::|||   ::::::||||||:::|||   ||||||||||||::
   10416 ATGTCCTCACCAAGAGTTGGTTTTATATTGCTTTTTTTCATTTTTGCCAT 10367

      17 rLeuAsnThrLeuLeuLeu 23
         |||:::   :::||||||
   10366 TCTACATCACATTTTATTA 10348

seq_name: gb_in2:AE003846

seq_documentation_block:
LOCUS       AE003846   298469 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386051 section 1
            of 2, complete sequence.
ACCESSION   AE003846 AE002804
VERSION     AE003846.2  GI:10728138
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 298469)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 298469)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7304344.
FEATURES             Location/Qualifiers
     source          1. .298469
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="4"
     mRNA            complement(1028. .>1306)
                     /gene="CG17469"
                     /product="CT38623"
                     /db_xref="FLYBASE:FBan0017469"
                     /db_xref="FLYBASE:FBgn0040845"
     gene            complement(<1028. .>1306)
                     /gene="CG17469"
                     /map="102E1-102E3"
                     /db_xref="FLYBASE:FBan0017469"
                     /db_xref="FLYBASE:FBgn0040845"
     CDS             complement(1028. .1192)
                     /gene="CG17469"
                     /note="CG17469 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0017469"
                     /db_xref="FLYBASE:FBgn0040845"
                     /protein_id="AAF59375.1"
                     /db_xref="GI:7304345"
                     /translation="MTESGIDLGFDMEFDLNINLLNDNDNMDFLPNVTENMEFYELKS
                     SSRCIRHNEM"
     mRNA            join(3595. .3608,3730. .3887,3945. .4042,4116. .4365,
                     4417. .4547,4611. .4779,6348. .6499,7372. .7501,7555. .>7671)
                     /gene="CG17471"
                     /product="CT38625"
                     /db_xref="FLYBASE:FBan0017471"
                     /db_xref="FLYBASE:FBgn0039924"
     gene            <3595. .>7671
                     /gene="CG17471"
                     /map="102E1-102E3"
                     /db_xref="FLYBASE:FBan0017471"
                     /db_xref="FLYBASE:FBgn0039924"
     CDS             join(3750. .3887,3945. .4042,4116. .4365,4417. .4547,
                     4611. .4779,6348. .6499,7372. .7501,7555. .7671)
                     /gene="CG17471"
                     /note="CG17471 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0017471"
                     /db_xref="FLYBASE:FBgn0039924"
                     /protein_id="AAF59376.2"
```

```
                /db_xref="GI:10728139"
                /translation="MEKKISSSSQPRILKKKHFRVKHQKVKLFRANEPILSVFMWGIN
                HTINELSHVNIPVMLLPDDFRAYSKIKVDNHLFNKENMPSHFKVKEYCPLVFRNLRER
                FGVDDVDYRESLTRSQPIQIDSSGKSGAQFYQSYDKFFIIKSLTSEEIERMHAFLKQY
                HPYVVERHGKTLLPQYLGMYRITVESVQYYFVVMRNVFSSHLTIHKKFDLKGSTVDRE
                ASEKELEKNLPTFKDNDFIKQKVKLDIGKEAKDKLMDTLSNDVDLLTKLHIMDYSLLV
                GVHDCVRAEEEALQGDNILTVGRSENSESEECDSGERFATNTTGIKIRELMYILNCIA
                EKREIYFIAIIDVLTQYGVKKQAAKAAKTVKYGSNVDGISTCDPEQYAKRFLDFMDKA
                IE"
mRNA            complement(join(7685. .7850,7901. .7982,8045. .8187,
                9658. .10140,10199. .10336))
                /gene="CG17467"
                /product="CT38621"
                /db_xref="FLYBASE:FBan0017467"
                /db_xref="FLYBASE:FBgn0022341"
gene            complement(<7685. .>10336)
                /gene="CG17467"
                /map="102E1-102E3"
                /db_xref="FLYBASE:FBan0017467"
                /db_xref="FLYBASE:FBgn0022341"
CDS             complement(join(7848. .7850,7901. .7982,8045. .8187,
                9658. .10140,10199. .10297))
                /gene="CG17467"
                /note="CG17467 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0017467"
                /db_xref="FLYBASE:FBgn0022341"
                /protein_id="AAF59377.1"
                /db_xref="GI:7304347"
                /translation="MKLDANIKSGLSNHVQYTLPDKKYQIHNQKVQENSDVEIVKSDS
                KPEEVKVATEELSKKFKFFETYSPAENKKRIFRMTPPREGVVMFPPPDSETNQQISTT
                LFNDNILQKTKTTSTILNKFREMEEQKMSDQQKKKNPKPLKCFTPPPEISHQFIRSDT
                EEESDSDYEQNSENDEESEINPSNSYYNDKALLEAQSVARAKQLRAKFEKWQNNEIEQ
                EIKEGRIDVYSQLISNESIESAKAIRERFENMKKSETAMENPSKTQIKRFV"
mRNA            join(18584. .18839,18906. .19163,25083. .25226,25280. .26049,
                26111. .26402,27005. .27124,27178. .27740)
                /gene="pho"
                /note="Nucleotide sequence of the Celera sequence differs
                from the published sequence for this transcript."
                /product="CT39329"
                /db_xref="FLYBASE:FBan0017743"
                /db_xref="FLYBASE:FBgn0002521"
gene            <18584. .>27740
                /gene="pho"
                /note="CG17743"
                /map="102E3-102E5"
                /db_xref="FLYBASE:FBan0017743"
                /db_xref="FLYBASE:FBgn0002521"
CDS             join(19069. .19163,25083. .25226,25280. .26049,26111. .26402,
                27005. .27124,27178. .27325)
                /gene="pho"
                /note="pho gene product; Nucleotide sequence of the Celera
                sequence differs from the published sequence for this
                transcript"
                /codon_start=1
                /db_xref="FLYBASE:FBan0017743"
                /db_xref="FLYBASE:FBgn0002521"
                /protein_id="AAF59378.1"
                /db_xref="GI:7304348"
                /translation="MAYERFGIILQSEQYDEDIGNTKVNQKMNEGKGNHYDLHRKNAF
                DRIIHSESKKGDNVINYNIHENDKIKAADNIFSSKLKMNPNMSYEMNINCFKNIGYGE
                NQETSKVLTNSLSNNDINTEESGVVDKNSPFLTLGTTILNSNGKSRRWEQKLVHIKTM
                EGEFSVTMWASGISDDEYSGSDQIVGASDLLKGKEEFGIDGFTSQQNKEYQKMESKFT
                NAQTLEMPHPISSVQIMDHLIKERGNLSQENNISERILSKTTLSFEEPILLPDSSSIE
                LVNETAAMTINNHRTLSNHTGNTGDLHALPSSVPFRIGLHEGQVNDCLSTISQSTHQD
                NTDSTGCGEMNLSEVTVSYTNDKKIACPHKGCNKHFRDSSAMRKHLHTHGPRVHVCAE
                CGKAFVESSKLKRHQLVHTGEKPFQCTFEGCGKRFSLDFNLRTHVRIHTGDRPFVCPF
                DACNKKFAQSTNLKSHILTHAKAKRNTSISGKSGCSNAESNSQSEDTSANYVKVELQD
                SVTENHVPFVVYAD"
mRNA            complement(join(33838. .34674,34722. .>35332))
                /gene="CG17461"
                /product="CT38613"
                /db_xref="FLYBASE:FBan0017461"
                /db_xref="FLYBASE:FBgn0039925"
gene            complement(<33838. .>35332)
                /gene="CG17461"
                /map="102E3-102E5"
                /db_xref="FLYBASE:FBan0017461"
                /db_xref="FLYBASE:FBgn0039925"
CDS             complement(join(33838. .34674,34722. .35255))
                /gene="CG17461"
                /note="CG17461 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0017461"
                /db_xref="FLYBASE:FBgn0039925"
                /protein_id="AAF59380.1"
                /db_xref="GI:7304350"
                /translation="MTTNVRYLALVTYLEIYNERIRDLLNKNENTNVINHFLKELPGI
                GVSVPTLTTQPVVNANQCYDWLHFGNKNRVTAATLMNKNSSRSHTIFTITLEQSPFLN
                SIGSDAFGGICRGKLSLVDLAGSERQRKTGAQGDRLKEASQINLSLSALGNVISSLVD
                GKAKHVPFRDSKLTRLLQYYFNKDSLGGNTKTLMISCISPTDIHYDETISTLRYASRA
                KNISNKPKINEDPKDARLRQYQNEILYLKRMLQESQQIINKNNDPNKIIKSPLKIIQH
                TNMNSTKNVQIIDLGRNCKASFKTNNSILTKPNFPLIQSKEEVQLQARSRIDLIKRSL
                IGGERIHDFELKEKHMARKYAAQRHLSAIAIALSRVKCEDRDLLQGHYATITQEIDIK
                NDYIRKCKEKIKMLEMEVSDLNSEFQLDREDYLDEIRNLGRQVKFHQQLFLKFSSNPG
                KFDRNW"
mRNA            complement(join(50956. .51220,52591. .52741))

alignment_scores:
            Quality:   60.00              Length:     20
              Ratio:  3.750                Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 55.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AE003846   ..

 Align seg 1/1  to: AE003846  from: 1  to: 298469

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||   ::::::||||||   ||||||   |||::::|||:::|||   ||
  109664 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 109713

      18 uAsnThrLeu 21
         ||||:::|||
  109714 AAATAGCCTA 109723

 seq_name: gb_htg7:AC021204

 seq_documentation_block:
 LOCUS       AC021204   158357 bp    DNA             HTG       17-AUG-2000
 DEFINITION  Homo sapiens chromosome X clone RP11-143G3, WORKING DRAFT SEQUENCE,
             15 unordered pieces.
 ACCESSION   AC021204
 VERSION     AC021204.6  GI:9838090
 KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 158357)
   AUTHORS   Waterston,R.H.
   TITLE     The sequence of Homo sapiens clone
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 158357)
   AUTHORS   Waterston,R.H.
   TITLE     Direct Submission
   JOURNAL   Submitted (15-JAN-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
 COMMENT     On Aug 17, 2000 this sequence version replaced gi:8954229.

             ---------------- Genome Center ----------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site:http://genome.wustl.edu/gsc/index.shtml
             -------------- Project Information --------------
```

```
          Center project name: H_NH0143G03
          ------------ Summary Statistics  --------------
          Sequencing vector: M13; 78%
          Sequencing vector: plasmid; 22%
          Chemistry: Dye-primer ET; 78% of reads
          Chemistry: Dye-terminator Big Dye; 22% of reads
          Assembly program: Phrap; version 0.990319
          Consensus quality: 145325 bases at least Q40
          Consensus quality: 149562 bases at least Q30
          Consensus quality: 152152 bases at least Q20
          Insert size: 175000; agarose-fp
          Insert size: 158286; sum-of-contigs
          Quality coverage: 3.43 in Q20 bases; agarose-fp
          Quality coverage: 3.70 in Q20 bases; sum-of-contigs
          ------------------------------------------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 15 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1539: contig of 1539 bp in length
          *     1540     1639: gap of unknown length
          *     1640     2801: contig of 1162 bp in length
          *     2802     2901: gap of unknown length
          *     2902     5208: contig of 2307 bp in length
          *     5209     5308: gap of unknown length
          *     5309     6887: contig of 1579 bp in length
          *     6888     6987: gap of unknown length
          *     6988     9583: contig of 2596 bp in length
          *     9584     9683: gap of unknown length
          *     9684    13291: contig of 3608 bp in length
          *    13292    13391: gap of unknown length
          *    13392    16412: contig of 3021 bp in length
          *    16413    16512: gap of unknown length
          *    16513    21055: contig of 4543 bp in length
          *    21056    21155: gap of unknown length
          *    21156    26232: contig of 5077 bp in length
          *    26233    26332: gap of unknown length
          *    26333    35392: contig of 9060 bp in length
          *    35393    35492: gap of unknown length
          *    35493    47474: contig of 11982 bp in length
          *    47475    47574: gap of unknown length
          *    47575    63901: contig of 16327 bp in length
          *    63902    64001: gap of unknown length
          *    64002    89634: contig of 25633 bp in length
          *    89635    89734: gap of unknown length
          *    89735   121750: contig of 32016 bp in length
          *   121751   121850: gap of unknown length
          *   121851   158357: contig of 36507 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-143G3"
BASE COUNT    50000 a  30000 c  29276 g  47671 t   1410 others
ORIGIN


alignment_scores:
            Quality:   59.50            Length:     22
              Ratio:  3.719              Gaps:      1
 Percent Similarity:  72.727   Percent Identity: 63.636

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC021204/rev   ..

 Align seg 1/1  to reverse of: AC021204  from: 1  to: 158357

       8 PheLeuLeuArg.............PhePheIlePheCysSerLeuAs 19
         ||||||||||||              |||   :::|||||||||||||||
    9417 TTTTTGCTCCGAGGGGGAGTCTCTTTTTTTGCCCTGTTTTGCTCCCTCAA 9368

      19 nThrLeuLeuLeuGly 24
         |:::||||||||||||
    9367 TGTTTTGCTGTTGGGG 9352

seq_name: gb_htg8:AC022211

seq_documentation_block:
LOCUS       AC022211   191151 bp    DNA             HTG       21-NOV-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-649A18, WORKING DRAFT
            SEQUENCE, 1 unordered pieces.
ACCESSION   AC022211
VERSION     AC022211.4  GI:11245626
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191151)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191151)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 21, 2000 this sequence version replaced gi:7631098.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0649A18
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 77%
            Sequencing vector: plasmid; 23%
            Chemistry: Dye-primer ET; 72% of reads
            Chemistry: Dye-terminator Big Dye; 28% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 190743 bases at least Q40
            Consensus quality: 191007 bases at least Q30
            Consensus quality: 191106 bases at least Q20
            Insert size: 204000; agarose-fp
            Insert size: 191151; sum-of-contigs
            Quality coverage: 7.54 in Q20 bases; agarose-fp
            Quality coverage: 8.05 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   191151: contig of 191151 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191151
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-649A18"
     misc_feature    1. .191151
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    48904 a  46522 c  47897 g  47828 t
ORIGIN
```

```
alignment_scores:
            Quality:   59.00              Length:      23
              Ratio:  3.105                Gaps:       0
 Percent Similarity:  82.609   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC022211   ..

 Align seg 1/1  to: AC022211  from: 1  to: 191151

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         :::|||   ::::::|||||||   |||||||   |||||||:::|||||||::
  105346 GTGAAGTGGTGCAATCTTGGCTCACTGCTCATTTTTTTTTTTTTTTGTAA 105395

      17 rLeuAsnThrLeuLeuLeu 23
         ::::|||   :::|||||||
  105396 CATAAATGGGGTTTTGCTA 105414

seq_name: gb_pr1:AC005002

seq_documentation_block:
LOCUS       AC005002   119841 bp    DNA             PRI       21-AUG-1999
DEFINITION  Homo sapiens clone DJ378P09, complete sequence.
ACCESSION   AC005002
VERSION     AC005002.2  GI:5757541
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119841)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119841)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 119841)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 21, 1999 this sequence version replaced gi:3212989.
FEATURES             Location/Qualifiers
     source          1. .119841
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ378P09"
BASE COUNT    37083 a  23671 c  21587 g  37500 t
ORIGIN


alignment_scores:
            Quality:   58.00              Length:      22
              Ratio:  3.222                Gaps:       0
 Percent Similarity:  81.818   Percent Identity: 54.545

alignment_block:
 US-09-215-435-177_COPY_1_24 x AC005002/rev   ..

 Align seg 1/1  to reverse of: AC005002  from: 1  to: 119841

       3 SerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAs 19
         |||::   |||   |||||||:::   |||||||:::||||||||||:::
   55073 TCCTGTATCCTCTTTTTTCTTTTTTTTTTTTTTTTTGTTTGTTCTGTTCC 55024

      19 nThrLeuLeuLeuGly 24
         |||::::::|||||||
   55023 AACTTTTATTTTAGGA 55008

seq_name: gb_pr9:HSJ612B15

seq_documentation_block:
LOCUS       HSJ612B15  127820 bp    DNA             PRI       30-SEP-2000
DEFINITION  Human DNA sequence from clone RP4-612B15 on chromosome 1p22.2-31.1.
            Contains the (possibly pseudo) gene for a novel protein similar to
            60S ribosomal protein L17 (RPL17), the gene for CGI-61, endophilin
            B1 and KIAA0491, ESTs, STSs, GSSs and two CpG islands, complete
            sequence.
ACCESSION   AL049597
VERSION     AL049597.29  GI:9663352
KEYWORDS    HTG; CGI-61; CpG island; endophilin B1; KIAA0491; RPL17.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127820)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 2, 2000 this sequence version replaced gi:9211554.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-612B15 is from the library RPCI-4 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP4-612B15.
FEATURES             Location/Qualifiers
     source          1. .127820
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p22.2-31.1"
                     /clone="RP4-612B15"
                     /clone_lib="RPCI-4"
     repeat_region   1. .476
                     /note="HERV9 repeat: matches 980. .1445 of consensus"
     repeat_region   477. .1327
                     /note="LTR12 repeat: matches 1. .1440 of consensus"
     repeat_region   1328. .4490
                     /note="HERV9 repeat: matches 1792. .4954 of consensus"
     repeat_region   4484. .4919
                     /note="Harlequin repeat: matches 840. .1271 of consensus"
     repeat_region   4909. .5191
                     /note="HERV17 repeat: matches 1255. .1540 of consensus"
     repeat_region   5166. .5302
                     /note="Harlequin repeat: matches 1584. .1720 of consensus"
     repeat_region   5297. .5937
```

```
                /note="HERV9 repeat: matches 7756. .8395 of consensus"
repeat_region   5949. .6034
                /note="pTR5 repeat: matches 1352. .1437 of consensus"
repeat_region   6035. .6641
                /note="LTR12 repeat: matches 1. .671 of consensus"
misc_feature    6295. .6797
                /note="match: GSS: Em:AQ746353"
repeat_region   6793. .6969
                /note="MIR repeat: matches 18. .163 of consensus"
repeat_region   6970. .7262
                /note="AluSg1 repeat: matches 2. .293 of consensus"
repeat_region   7263. .7350
                /note="MIR repeat: matches 163. .258 of consensus"
repeat_region   7355. .7495
                /note="L2 repeat: matches 2559. .2705 of consensus"
misc_feature    complement(7658. .8075)
                /note="match: GSS: Em:AQ138017"
repeat_region   8194. .8491
                /note="L2 repeat: matches 1995. .2305 of consensus"
repeat_region   9635. .9924
                /note="AluSg/x repeat: matches 132. .291 of consensus"
repeat_region   9925. .10223
                /note="AluSc repeat: matches 1. .299 of consensus"
repeat_region   10252. .10943
                /note="346 copies 2 mer tt 58% conserved"
misc_feature    10502. .10783
                /note="Tandem repeat. Single clone region. Restriction
                digest (BAMHI and HINDIII) and pUC partners confirm
                assembly."
repeat_region   10966. .11254
                /note="AluSq repeat: matches 1. .288 of consensus"
repeat_region   11637. .11929
                /note="AluSp repeat: matches 1. .295 of consensus"
repeat_region   12129. .12503
                /note="LTR16C repeat: matches 1. .387 of consensus"
repeat_region   12565. .13003
                /note="LTR37B repeat: matches 9. .468 of consensus"
repeat_region   13019. .13306
                /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region   13727. .14008
                /note="AluSq repeat: matches 1. .284 of consensus"
repeat_region   14022. .14337
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   14400. .14449
                /note="25 copies 2 mer aa 76% conserved"
repeat_region   15656. .15771
                /note="L1MA10 repeat: matches 6212. .6322 of consensus"
repeat_region   15767. .15944
                /note="L1MC1 repeat: matches 6154. .6332 of consensus"
repeat_region   15966. .16245
                /note="L1MB8 repeat: matches 5781. .6078 of consensus"
repeat_region   16261. .16313
                /note="L1MA10 repeat: matches 6233. .6285 of consensus"
repeat_region   16346. .16538
                /note="L1MB8 repeat: matches 5595. .5791 of consensus"
repeat_region   16539. .16830
                /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region   16831. .17074
                /note="L1MB8 repeat: matches 5364. .5595 of consensus"
repeat_region   17111. .17158
                /note="24 copies 2 mer tg 81% conserved"
repeat_region   17172. .17377
                /note="L1MB8 repeat: matches 5126. .5349 of consensus"
repeat_region   17474. .17778
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   18332. .18584
                /note="AluJb repeat: matches 14. .265 of consensus"
repeat_region   18665. .18769
                /note="MIR repeat: matches 103. .212 of consensus"
repeat_region   19024. .19170
                /note="L2 repeat: matches 2479. .2624 of consensus"
repeat_region   19771. .20967
                /note="L1MB2 repeat: matches 4967. .6168 of consensus"
repeat_region   20970. .21198
                /note="MSTB repeat: matches 192. .426 of consensus"
repeat_region   21203. .21345
                /note="AluY repeat: matches 156. .298 of consensus"
repeat_region   21354. .21536
                /note="MSTB repeat: matches 1. .189 of consensus"
repeat_region   21537. .21966
                /note="L1MB2 repeat: matches 4538. .4969 of consensus"
repeat_region   21984. .22039
                /note="L1 repeat: matches 4545. .4599 of consensus"
repeat_region   22029. .22753
                /note="L1MB2 repeat: matches 3821. .4538 of consensus"
misc_feature    22746. .22957
                /note="match: GSS: Em:AQ279252"
misc_feature    complement(22749. .23381)
                /note="match: GSS: Em:AZ040837"
misc_feature    22751. .22892
                /note="match: GSS: Em:AQ265038"
misc_feature    22752. .23055
                /note="match: GSS: Em:AQ543943"
misc_feature    22761. .23066
                /note="match: STS: Em:G06064"
misc_feature    22764. .23165
                /note="match: GSS: Em:AQ194597"
gene            complement(22792. .23346)
                /gene="dJ612B15.1"
CDS             complement(22792. .23346)
                /gene="dJ612B15.1"
                /note="probably a pseudogene
                match: proteins: Sw:P35266 Tr:O64803 Sw:P35267 Sw:O48557
                Sw:P18621 Tr:Q9W3W8 Sw:P54033 Sw:O59423 Sw:O59794
                Tr:Q9Z322 Sw:P05740 Sw:P46990 Sw:O28359 Sw:P37380
                Sw:O26115 Sw:P24049 Tr:O14339"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ612B15.1 (novel protein similar to 60S
                ribosomal protein L17 (RPL17))"
                /protein_id="CAC10392.1"
                /db_xref="GI:10443242"
                /translation="MVRYSLDPENPMKSCKSRGSNLRVHFKDTRETAQAIKGMHIRKA
                TKYLKDVTLQKQCIPFRRYNGGVGKCAQAKQWGWTQGRWPKKSAEFLLHMFKNAEGNT
                ELKGLDVDSLVIEYIQVNKAPKMLRRTYRAHGRINSYMSSPWHIEMILTEKEQIVPKP
                EKEVAQKKKISQKKLKKQKLMARK"
misc_feature    complement(22844. .23381)
                /note="match: GSS: Em:AZ040921"
misc_feature    complement(23031. .23133)
                /gene="dJ612B15.1"
                /note="match: GSS: Em:AQ808942"
misc_feature    23130. .23276
                /note="match: GSS: Em:AQ793153"
repeat_region   23381. .23598
                /note="L1MB2 repeat: matches 3608. .3834 of consensus"
repeat_region   23599. .23679
                /note="AluSc repeat: matches 228. .306 of consensus"
repeat_region   23683. .24713
                /note="L1MB2 repeat: matches 2580. .3637 of consensus"
repeat_region   24714. .24993
                /note="AluY repeat: matches 17. .302 of consensus"
repeat_region   24994. .25264
                /note="L1MB2 repeat: matches 2315. .2580 of consensus"

alignment_scores:
            Quality:   58.00              Length:     24
              Ratio:  3.412                Gaps:      0
 Percent Similarity:  70.833   Percent Identity: 54.167

alignment_block:
 US-09-215-435-177_COPY_1_24 x HSJ612B15/rev   ..

 Align seg 1/1  to reverse of: HSJ612B15  from: 1  to: 127820

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         :::|||      :::|||   ||||||||||||:::|||   :::|||
```

```
9378 TTAAAACTACTTCAACTCAAATTCTTGCTCAGGATGTTCTCTATTTGTTT 9329

  17 rLeuAsnThrLeuLeuLeuGly 24
       ||||||||||   |||   |||
9328 CCTGAACACCAGATTATCTGGG 9307
```

OM of: US-09-215-435-177_COPY_1_24 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455	+	120.00	322.51	5.4e-10	267	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569	+	120.00	322.38	5.5e-10	271	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612	+	120.00	316.42	1.2e-09	526	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738	+	120.00	316.20	1.2e-09	539	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13059	-	54.00	111.77	296.28	9813		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF22305	+	54.00	69.51	6.5e+04	108213		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34882	+	52.00	124.72	56.25	1272		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF22295	-	52.00	87.46	6.7e+03	80450		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC21845	+	51.00	136.57	12.31	252	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72760	-	50.50	121.90	80.77	1109		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF16129	-	50.50	115.35	187.14	2299		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72785	-	50.50	115.35	187.14	2299		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV43009	+	50.00	113.97	223.44	2307		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAZ96363	+	50.00	112.27	277.82	2787		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV52185	+	50.00	96.24	2.2e+03	16593		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32057	+	49.50	123.31	67.44	702	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA99912	+	49.00	127.26	40.62	389	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX29493	-	49.00	120.28	99.44	846	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX19930	-	49.00	119.29	112.97	945	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ70890	+	49.00	108.76	435.77	3049		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46410	-	49.00	106.60	575.13	3879		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ14000	-	49.00	105.88	630.83	4203		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF44629	+	49.00	104.35	767.57	4983		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV69920	-	49.00	101.49	1.1e+03	6850		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74402	-	49.00	97.22	1.9e+03	11014		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ34585	+	48.50	112.78	260.20	1677		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC14840	+	48.00	128.25	35.78	258	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA44260	-	48.00	126.63	44.05	309	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC56935	+	48.00	125.38	51.69	355	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC56987	+	48.00	124.18	60.34	406	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC02642	+	48.00	123.17	68.64	454	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT64942	+	48.00	121.71	82.75	534	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA97369	+	48.00	112.25	278.60	1531		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC46992	+	48.00	111.86	292.69	1598		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT93210	+	48.00	111.55	304.55	1654		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT64943	+	48.00	111.55	304.55	1654		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08500	+	48.00	111.55	304.55	1654		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ58929	+	48.00	111.55	304.55	1654		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF59257	+	48.00	111.55	304.55	1654		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT67494	-	48.00	109.38	402.32	2106		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT68246	-	48.00	108.24	465.69	239		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT33677	-	48.00	108.19	469.06	240		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08038	-	48.00	104.91	714.16	346		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV07010	+	48.00	103.98	804.19	384		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13453	+	48.00	101.43	1.1e+03	510		

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455

seq_documentation_block:
ID AAC00455 standard; cDNA; 267 BP.
XX
AC AAC00455;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 453.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00449.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 453; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 267 BP; 74 A; 57 C; 52 G; 84 T; 0 other;

alignment_scores:
Quality: 120.00 Length: 24
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-177_COPY_1_24 x AAC00455 ..

Align seg 1/1 to: AAC00455 from: 1 to: 267

1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
||||||||||||||||||||||||||||||||||||||||||||||||||

```
      31 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 80

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      81 ATTGAATACCCTGTTATTGGGT 102
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569

seq_documentation_block:
ID AAX51569 standard; cDNA; 271 BP.
XX
AC AAX51569;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:148.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
DR P-PSDB; AAY12791.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Claim 1; Page 253-254; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 271 BP; 75 A; 57 C; 55 G; 84 T; 0 other;

alignment_scores:
Quality: 120.00 Length: 24
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-215-435-177_COPY_1_24 x AAX51569

Align seg 1/1 to: AAX51569 from: 1 to: 271

```
       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGT 103
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612

seq_documentation_block:
ID AAX97612 standard; DNA; 526 BP.
XX
AC AAX97612;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 76.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; AAY35928.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 210; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 526 BP; 162 A; 100 C; 99 G; 165 T; 0 other;

alignment_scores:
Quality: 120.00 Length: 24
Ratio: 5.000 Gaps: 0

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAX97612    .;

 Align seg 1/1  to: AAX97612  from: 1  to: 526

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      72 ATTGAATACCCTGTTATTGGGT 93

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738

seq_documentation_block:
ID   AAX97738 standard; DNA; 539 BP.
XX
AC   AAX97738;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 303.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY36054.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 376; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 539 BP; 165 A; 103 C; 97 G; 162 T; 12 other;

alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAX97738    ..

 Align seg 1/1  to: AAX97738  from: 1  to: 539

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGT 103

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13059

seq_documentation_block:
ID   AAX13059 standard; DNA; 9813 BP.
XX
AC   AAX13059;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:122.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 734-739; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 9813 BP; 3233 A; 1698 C; 2161 G; 2719 T; 2 other;
```

```
alignment_scores:
            Quality:   54.00              Length:     21
              Ratio:  3.600                Gaps:      0
 Percent Similarity:  71.429   Percent Identity: 52.381

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAX13059/rev   ..

 Align seg 1/1  to reverse of: AAX13059  from: 1  to: 9813

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         |||      |||   :::   |||:::   ||||||||||:::||||||||
     408 ATGGTTATGGCCTTCATGACCTTCTTCGCCAGATTCTTCCTCTTCTGTTC 359

      17 rLeuAsnThrLeu 21
         ||||:::   |||
     358 CTTGCATCAATTG 346

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF22305

seq_documentation_block:
ID   AAF22305 standard; DNA; 1082138 BP.
XX
AC   AAF22305;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Arabidopsis thaliana chromosome 4 centromere.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells
XX
PS   Claim 68; Page 977-1388; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

alignment_scores:
            Quality:   54.00              Length:     16
              Ratio:  3.857                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 62.500

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAF22305   ..

 Align seg 1/1  to: AAF22305  from: 1  to: 1082138

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSer 17
         |||:::::::||||||:::|||   :::   ||||||||||||||||||
  526358 AAAGATTCAAAACTATCATTTAAAATTTACTTCTTCATTTTTTGCTCT 526405

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34882

seq_documentation_block:
ID   AAC34882 standard; DNA; 1272 BP.
XX
AC   AAC34882;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 8233.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
```

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality:	52.00	Length:	19
Ratio:	3.467	Gaps:	0
Percent Similarity:	78.947	Percent Identity:	52.632

```
alignment_block:
 US-09-215-435-177_COPY_1_24 x AAC34882    ..

 Align seg 1/1  to: AAC34882  from: 1  to: 1272

       5 LysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLe 21
         |||:::|||||||:::   ||||||:::::::|||   ||||||   |||
     135 AAAATGGGTTTTATAACTCGATTTTTAGTTTTCATGTCACTCTTCACCAG 184

      21 uLeuLeu 23
         |||:::
     185 TTTGGTT 191

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF22295

seq_documentation_block:
ID   AAF22295 standard; DNA; 80450 BP.
XX
AC   AAF22295;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #18.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 668-686; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 80450 BP; 25014 A; 14662 C; 14629 G; 26145 T; 0 other;


alignment_scores:
            Quality:   52.00              Length:     22
              Ratio:  2.889                Gaps:      1
 Percent Similarity:  81.818   Percent Identity: 50.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAF22295/rev    ..

 Align seg 1/1  to reverse of: AAF22295  from: 1  to: 80450

       5 LysLeuGlyPheLeuLeuArg......PhePheIlePheCysSerLeuAs 19
         |||||||||||||:::::::::      |||:::   ||||||:::|||::
     31464 AAGCTTGGCTTCATGGTTAAGAAGTCGTTCATGGACTTCTGTGGTTTGAG 31415

      19 nThrLeuLeuLeuGly 24
         :   |||:::|||||||
     31414 GAGGCTTATCCTTGGC 31399

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC21845

seq_documentation_block:
ID   AAC21845 standard; cDNA; 252 BP.
XX
AC   AAC21845;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 25920.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 25920; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 252 BP; 72 A; 49 C; 64 G; 66 T; 1 other;


alignment_scores:
            Quality:   51.00              Length:     16
              Ratio:  3.643                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 56.250

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAC21845    ..

 Align seg 1/1  to: AAC21845  from: 1  to: 252

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeu 22
         ||||||||||||:::|||||||   :::|||:::   |||:::::|||
```

```
    161 GGCTTTCTCTTGAAGTTTTTTGCCATATGTACACAGAATGCATTTCTG 208

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72760

seq_documentation_block:
ID   AAF72760 standard; cDNA; 1109 BP.
XX
AC   AAF72760;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human prostate cancer antigen coding sequence #20.
XX
KW   Immunosuppressive; nootropic; neuroprotective; antiviral; vulnerary;
KW   anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant;
KW   immune disorder; cardiovascular disorder; neurological disease;
KW   infection; cancer; cytostatic; antiarthritic; antirheumatic;
KW   antiasthmatic; anticonvulsant; vasotropic; vulnerary; human;
KW   secreted protein; prostate cancer antigen; ss.
XX
OS   Homo sapiens.
XX
PN   WO200107476-A1.
XX
PD   01-FEB-2001.
XX
PF   20-JUL-2000; 2000WO-US19666.
XX
PR   21-JUL-1999;   99US-0144972.
PR   13-AUG-1999;   99US-0148681.
PR   17-AUG-1999;   99US-0149173.
PR   06-OCT-1999;   99US-0158004.
PR   05-APR-2000; 2000US-0194689.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Birse C;
XX
DR   WPI; 2001-138554/14.
DR   P-PSDB; AAB80292.
XX
PT   New nucleic acid molecule encoding human secreted prostate cancer
PT   antigens, useful for the diagnosis and treatment of disorders such as
PT   cancer, leukemia and autoimmune disease -
XX
PS   Claim 1; Page 369; 433pp; English.
XX
CC   The present invention relates to human secreted prostate cancer antigen
CC   coding sequences (AAF72741-AAF72789) and proteins (AAB80273-AAB80321).
CC   The coding sequences and proteins of the present invention are useful for
CC   preventing, treating or ameliorating a medical condition; and for the
CC   diagnosis and treatment of diseases and disorders. Diseases and disorders
CC   that can be diagnosed and treated include (auto)immune diseases (e.g.
CC   graft versus host disease and rheumatoid arthritis), inflammatory and
CC   allergic disorders (e.g. asthma), hyperproliferative disorders (e.g.
CC   cancers and leukemias), cardiovascular disorders (e.g. heart attacks and
CC   arrythmias), cerebrovascular disorders (e.g. stroke), arterial occlusive
CC   disorders (e.g. arteriosclerosis), angiogenesis related disorders (e.g.
CC   retinopathy and keloid scars), ocular disorders (e.g. glaucoma),
CC   neurological disorders (e.g. Alzheimer's, Parkinson's disease, epilepsy
CC   and Creutzfeld-Jakob disease) and infections caused by bacteria, fungi,
CC   viruses or parasites. They may also be useful for wound healing,
CC   epithelial cell proliferation, supporting cell culture, tissue
CC   regeneration, birth control and as a food additive or preservative.
CC   The coding sequences can be used to generate fusion proteins by linking
CC   the coding sequences to the human immunoglobulin G Fc portion coding
CC   sequence (AAF72732) for increasing the stability of the fusion protein
CC   as compared to the human protein only.
XX
SQ   Sequence 1109 BP; 346 A; 211 C; 198 G; 354 T; 0 other;

alignment_scores:
            Quality:   50.50              Length:     23
              Ratio:  2.658                Gaps:      1
 Percent Similarity:  82.609   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAF72760/rev   ..

 Align seg 1/1  to reverse of: AAF72760  from: 1  to: 1109

       1 MetLysSerAlaLysLeu...GlyPheLeuLeuArgPhePheIlePheCy 16
         |||:::|||   |||:::      :::||||||||||||:::|||   ||
     434 ATGAGTTCAGATAAAATGTTCAAATATTTGCTTAGATTTTTAATCGGTTG 385

      16 sSerLeuAsnThrLeuLeu 22
         ||||::::::||||||:::
     384 CTCTATGAGAACCCTCATA 366

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF16129

seq_documentation_block:
ID   AAF16129 standard; cDNA; 2299 BP.
XX
AC   AAF16129;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:564.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB56926.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 1022-1023; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
```

```
CC   invention.
XX
SQ   Sequence 2299 BP; 666 A; 437 C; 417 G; 774 T; 5 other;


alignment_scores:
            Quality:   50.50              Length:     23
              Ratio:  2.658                Gaps:      1
 Percent Similarity:  82.609   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAF16129/rev   ..

 Align seg 1/1  to reverse of: AAF16129  from: 1  to: 2299

       1 MetLysSerAlaLysLeu...GlyPheLeuLeuArgPhePheIlePheCy 16
         |||:::|||   |||:::      :::|||||||||||||:::|||   ||
    1652 ATGAGTTCAGATAAAATGTTCAAATATTTGCTTAGATTTTTAATCGGTTG 1603

      16 sSerLeuAsnThrLeuLeu 22
         ||||::::::||||||:::
    1602 CTCTATGAGAACCCTCATA 1584

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF72785

 seq_documentation_block:
 ID   AAF72785 standard; cDNA; 2299 BP.
 XX
 AC   AAF72785;
 XX
 DT   24-APR-2001  (first entry)
 XX
 DE   Human prostate cancer antigen coding sequence #45.
 XX
 KW   Immunosuppressive; nootropic; neuroprotective; antiviral; vulnerary;
 KW   anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant;
 KW   immune disorder; cardiovascular disorder; neurological disease;
 KW   infection; cancer; cytostatic; antiarthritic; antirheumatic;
 KW   antiasthmatic; anticonvulsant; vasotropic; vulnerary; human;
 KW   secreted protein; prostate cancer antigen; ss.
 XX
 OS   Homo sapiens.
 XX
 PN   WO200107476-A1.
 XX
 PD   01-FEB-2001.
 XX
 PF   20-JUL-2000; 2000WO-US19666.
 XX
 PR   21-JUL-1999;   99US-0144972.
 PR   13-AUG-1999;   99US-0148681.
 PR   17-AUG-1999;   99US-0149173.
 PR   06-OCT-1999;   99US-0158004.
 PR   05-APR-2000; 2000US-0194689.
 XX
 PA   (HUMA-) HUMAN GENOME SCI INC.
 PA   (ROSE/) ROSEN C A.
 XX
 PI   Rosen CA,  Birse C;
 XX
 DR   WPI; 2001-138554/14.
 DR   P-PSDB; AAB80317.
 XX
 PT   New nucleic acid molecule encoding human secreted prostate cancer
 PT   antigens, useful for the diagnosis and treatment of disorders such as
 PT   cancer, leukemia and autoimmune disease -
 XX
 PS   Claim 1; Page 382; 433pp; English.
 XX
 CC   The present invention relates to human secreted prostate cancer antigen
 CC   coding sequences (AAF72741-AAF72789) and proteins (AAB80273-AAB80321).
 CC   The coding sequences and proteins of the present invention are useful for
 CC   preventing, treating or ameliorating a medical condition; and for the
 CC   diagnosis and treatment of diseases and disorders. Diseases and disorders
 CC   that can be diagnosed and treated include (auto)immune diseases (e.g.
 CC   graft versus host disease and rheumatoid arthritis), inflammatory and
 CC   allergic disorders (e.g. asthma), hyperproliferative disorders (e.g.
 CC   cancers and leukemias), cardiovascular disorders (e.g. heart attacks and
 CC   arrythmias), cerebrovascular disorders (e.g. stroke), arterial occlusive
 CC   disorders (e.g. arteriosclerosis), angiogenesis related disorders (e.g.
 CC   retinopathy and keloid scars), ocular disorders (e.g. glaucoma),
 CC   neurological disorders (e.g. Alzheimer's, Parkinson's disease, epilepsy
 CC   and Creutzfeld-Jakob disease) and infections caused by bacteria, fungi,
 CC   viruses or parasites. They may also be useful for wound healing,
 CC   epithelial cell proliferation, supporting cell culture, tissue
 CC   regeneration, birth control and as a food additive or preservative.
 CC   The coding sequences can be used to generate fusion proteins by linking
 CC   the coding sequences to the human immunoglobulin G Fc portion coding
 CC   sequence (AAF72732) for increasing the stability of the fusion protein
 CC   as compared to the human protein only.
 XX
 SQ   Sequence 2299 BP; 666 A; 437 C; 417 G; 774 T; 5 other;


alignment_scores:
            Quality:   50.50              Length:     23
              Ratio:  2.658                Gaps:      1
 Percent Similarity:  82.609   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAF72785/rev   ..

 Align seg 1/1  to reverse of: AAF72785  from: 1  to: 2299

       1 MetLysSerAlaLysLeu...GlyPheLeuLeuArgPhePheIlePheCy 16
         |||:::|||   |||:::      :::|||||||||||||:::|||   ||
    1652 ATGAGTTCAGATAAAATGTTCAAATATTTGCTTAGATTTTTAATCGGTTG 1603

      16 sSerLeuAsnThrLeuLeu 22
         ||||::::::||||||:::
    1602 CTCTATGAGAACCCTCATA 1584

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV43009

 seq_documentation_block:
 ID   AAV43009 standard; DNA; 2307 BP.
 XX
 AC   AAV43009;
 XX
 DT   09-NOV-1998  (first entry)
 XX
 DE   Streptococcus pneumoniae polypeptide coding region.
 XX
 KW   Polypeptide; ORF; open reading frame; infection; bacterial;
 KW   streptococcal; bacteremia; diagnosis; prophylaxis; ds.
 XX
 OS   Streptococcus pneumoniae.
 XX
 FH   Key             Location/Qualifiers
 FT   CDS             1878..2306
 FT                   /*tag= a
 FT                   /note= "polypeptide"
 XX
 PN   WO9823631-A1.
 XX
 PD   04-JUN-1998.
 XX
 PF   24-NOV-1997;   97WO-US21976.
 XX
 PR   27-NOV-1996;   96US-0031879.
 XX
 PA   (SMIK ) SMITHKLINE BEECHAM CORP.
 PA   (SMIK ) SMITHKLINE BEECHAM PLC.
 XX
 PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
 PI   Reid RH,  Zarfos PN;
```

```
XX
DR   WPI; 1998-322654/28.
DR   P-PSDB; AAW62734.
XX
PT   Streptococcus pneumoniae polynucleotides - useful for developing
PT   products for diagnosis, prevention and treatment of infections e.g.
PT   pneumonia, bacteremia, meningitis or endocarditis
XX
PS   Claim 1; Page 127; 181pp; English.
XX
CC   The sequence is that of a Streptococcal polypeptide coding region.
CC   The polypeptide can potentially be used for the diagnosis and
CC   prevention of bacterial infections, especially SP infection.
CC   It may be used for the treatment of diseases such as otitis media,
CC   conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC   empyema, endocarditis or infection of the cerebrospinal fluid.
XX
SQ   Sequence 2307 BP; 742 A; 362 C; 509 G; 694 T; 0 other;


alignment_scores:
            Quality:   50.00              Length:     17
              Ratio:  3.571                Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 47.059

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAV43009   ..

 Align seg 1/1  to: AAV43009  from: 1  to: 2307

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         |||||||||:::::   ::::::||||||   ||||||   :::|||::
    1610 GGATTTTTGATGGAAGTGATGGTGTTTTGCGTTTTAAACAGAATTTTAAT 1659

      23 u 23
         :
    1660 G 1660

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAZ96363

seq_documentation_block:
ID   AAZ96363 standard; DNA; 2787 BP.
XX
AC   AAZ96363;
XX
DT   10-APR-2000  (first entry)
XX
DE   S. pneumoniae derived DNA from ORF #191.
XX
KW   Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW   bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9806734-A1.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US14436.
XX
PR   16-AUG-1996;   96US-0024022.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-159452/14.
DR   P-PSDB; AAY86028, AAY86029, AAY86030.
XX
PT   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
PS   Claim 4; Page 222-223; 640pp; English.
XX
CC   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC   AAY85792-Y86182). The DNA, vectors and host cells described in the
CC   method of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
SQ   Sequence 2787 BP; 883 A; 449 C; 607 G; 848 T; 0 other;


alignment_scores:
            Quality:   50.00              Length:     17
              Ratio:  3.571                Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 47.059

alignment_block:
 US-09-215-435-177_COPY_1_24 x AAZ96363   ..

 Align seg 1/1  to: AAZ96363  from: 1  to: 2787

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         |||||||||:::::   ::::::||||||   ||||||   :::|||::
    2090 GGATTTTTGATGGAAGTGATGGTGTTTTGCGTTTTAAACAGAATTTTAAT 2139

      23 u 23
         :
    2140 G 2140

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV52185

seq_documentation_block:
ID   AAV52185 standard; DNA; 16593 BP.
XX
AC   AAV52185;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:52.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
```

XX
PS Claim 1; Page 466-475; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 16593 BP; 4970 A; 2931 C; 3945 G; 4746 T; 1 other;

alignment_scores:
Quality: 50.00 Length: 17
Ratio: 3.571 Gaps: 0
Percent Similarity: 82.353 Percent Identity: 47.059

alignment_block:
US-09-215-435-177_COPY_1_24 x AAV52185 ..

Align seg 1/1 to: AAV52185 from: 1 to: 16593

```
      7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
        ||||||||||::::::   ::::::||||||   ||||||   :::|||::
  11958 GGATTTTTGATGGAAGTGATGGTGTTTTGCGTTTTAAACAGAATTTTAAT 12007

     23 u 23
        :
  12008 G 12008
```

```
OM of: US-09-215-435-177_COPY_1_24 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  3:58 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000


score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-943-731-2 +   50.00   91.03  2.0e+03   38682  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-298-367B-2 -   49.00  104.48  360.55     6850  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-991-426-3 +   48.00  115.02  93.33      1654  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-143-470-3 +   48.00  115.02  93.33      1654  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-396-479B-7 -   48.00  111.50  146.64     2406  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-818-823-7 -   48.00  111.50  146.64     2406  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-219-932-5 +   48.00  107.10  257.75     3841  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-623-906A-3 -   47.00  128.93  15.68      279  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-949-246-3 +   47.00  112.38  130.90     1622  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-635-561A-7 +   46.50  132.63  9.75       162  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-248-335-57 -   46.00  114.49  99.88      960  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-329-796-2 -   46.00  112.54  128.22     1181  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-415-593-44 -   46.00  111.27  150.93     1352  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-123-343A-6 -   46.00  101.31  541.36     3900  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-257-073-6 -   45.50  113.02  120.59     966  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-602-359A-27 -   45.00  112.34  131.62     894  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-899-324-32 +   45.00  102.19  483.88     2632  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-329-892B-32 +   45.00  102.19  483.88     2632  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-617-860B-35 -   45.00   99.92  647.20     3350  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-605-106-7 -   45.00   95.32  1.2e+03    5467  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-394-189B-1 -   45.00   92.16  1.8e+03    7653  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-258-287B-1 -   45.00   92.16  1.8e+03    7653  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-368-704C-1 -   45.00   92.16  1.8e+03    7653  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-05701-18 -   45.00   92.16  1.8e+03   7653
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-05705-1 -   45.00   92.16  1.8e+03   7653
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-285-440-23 -   44.00  120.27  47.63      285  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-630-349-23 -   44.00  120.27  47.63      285  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-385-982-229 +   44.00  113.11  119.22     610  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-408-095-26 +   44.00  108.99  202.34     946  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-100-391-5 -   44.00  107.00  261.17     1169  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-100-391-11 -   44.00  105.98  297.40     1302  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-08177-18 -   44.00  102.54  462.84    1879
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-100-391-9 -   44.00   99.33  698.35     2643  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-248-466B-2 -   44.00   98.91  736.75     2763  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-972-927-2 +   44.00   86.88  3.4e+03    9936  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-966-958-1 -   44.00   86.14  3.8e+03   10754  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-09-215-817-1 -   44.00   86.14  3.8e+03   10754  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-342-353-1 -   44.00   86.14  3.8e+03   10754  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-721-925A-4 +   43.00  116.24  79.83      324  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-714-918-47 +   43.00  112.49  129.20     483  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-265-315-47 +   43.00  112.49  129.20     483
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-265-315-47 +   43.00  112.49  129.20     483
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-266-417-47 +   43.00  112.49  129.20     483
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-955-629C-1 -   43.00  107.67  239.54     806
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-06661-1 -   43.00  100.85  574.89     16

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-943-731-2

seq_documentation_block:
; Sequence 2, Application US/08943731
; Patent No. 6265157
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL.
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  38682 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-2


alignment_scores:
            Quality:   50.00              Length:     20
              Ratio:  3.333                Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 60.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-943-731-2   ..
```

```
Align seg 1/1  to: US-08-943-731-2  from: 1  to: 38682

      5 LysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLe 21
        ||||||   |||||||||   ::::::::::|||   |||||||||
  17744 AAATTATTTTTTTTACTCCCTCTTCTTTTGTTCTTTTCATTAAACAGGGC 17793

     21 uLeuLeuGly 24
        |||||||||
  17794 TCTGCTGGGC 17803

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-298-367B-2

seq_documentation_block:
; Sequence 2, Application US/09298367B
; Patent No. 6180112
; GENERAL INFORMATION:
;  APPLICANT:  Highlander, Sarah K.
;  APPLICANT:  Federova, Natalie D.
;  TITLE OF INVENTION:  PASTEURELLA HAEMOLYTICA VACCINE
;  FILE REFERENCE:  BCM-03728
;  CURRENT APPLICATION NUMBER: US/09/298,367B
;  CURRENT FILING DATE:  1999-04-22
;  PRIOR APPLICATION NUMBER:  08/834,455
;  PRIOR FILING DATE:  1997-04-15
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 6850
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
US-09-298-367B-2


alignment_scores:
            Quality:   49.00            Length:     14
              Ratio:  4.455              Gaps:      0
 Percent Similarity:  78.571   Percent Identity: 64.286

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-298-367B-2/rev   ..

 Align seg 1/1  to reverse of: US-09-298-367B-2  from: 1  to: 6850

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   |||   :::|||||||||||||||||   :::
     700 TTTTTGTTGTTTTTTGTTGTTTTTTGTAGCTTAAATTGTATT 659

 seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-991-426-3

 seq_documentation_block:
 ; Sequence 3, Application US/08991426
 ; Patent No. 6013257
 ;  GENERAL INFORMATION:
 ;     APPLICANT:  Pan, Yang
 ;     TITLE OF INVENTION:  NEUROTACTIN AND USES THEREFOR
 ;     NUMBER OF SEQUENCES:  18
 ;     CORRESPONDENCE ADDRESS:
 ;       ADDRESSEE:  Fish & Richardson P.C.
 ;       STREET:  225 Franklin Street
 ;       CITY:  Boston
 ;       STATE:  MA
 ;       COUNTRY:  US
 ;       ZIP:  02110-2804
 ;     COMPUTER READABLE FORM:
 ;       MEDIUM TYPE:  Diskette
 ;       COMPUTER:  IBM Compatible
 ;       OPERATING SYSTEM:  Windows95
 ;       SOFTWARE:  FastSEQ for Windows Version 2.0
 ;     CURRENT APPLICATION DATA:
 ;       APPLICATION NUMBER:  US/08/991,426
 ;       FILING DATE:  16-DEC-1997
 ;     PRIOR APPLICATION DATA:
 ;       APPLICATION NUMBER:  08/851,160
 ;       FILING DATE:  05-MAY-1997
 ;     PRIOR APPLICATION DATA:
 ;       APPLICATION NUMBER:  08/643,798
 ;       FILING DATE:  07-MAY-1996
 ;     ATTORNEY/AGENT INFORMATION:
 ;       NAME:  Meiklejohn, Ph.D., Anita L.
 ;       REGISTRATION NUMBER:  35,283
 ;       REFERENCE/DOCKET NUMBER:  09404/043001
 ;     TELECOMMUNICATION INFORMATION:
 ;       TELEPHONE:  617/542-5070
 ;       TELEFAX:  617/542-8906
 ;       TELEX:  200154
 ;  INFORMATION FOR SEQ ID NO:  3:
 ;     SEQUENCE CHARACTERISTICS:
 ;       LENGTH:  1654 base pairs
 ;       TYPE:  nucleic acid
 ;       STRANDEDNESS:  single
 ;       TOPOLOGY:  linear
 ;     MOLECULE TYPE:  cDNA
 ;     FEATURE:
 ;       NAME/KEY:  Coding Sequence
 ;       LOCATION:  86...1276
 US-08-991-426-3


 alignment_scores:
             Quality:   48.00            Length:     24
               Ratio:  2.824              Gaps:      0
  Percent Similarity:  70.833   Percent Identity: 45.833

 alignment_block:
  US-09-215-435-177_COPY_1_24 x US-08-991-426-3   ..

  Align seg 1/1  to: US-08-991-426-3  from: 1  to: 1654

        1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
          |||         :::|||:::::::||||||||||:::      ||||||
       86 ATGGCTCCGATATCTCTGTCGTGGCTGCTCCGCTTGGCCACCTTCTGCCA 135

       17 rLeuAsnThrLeuLeuLeuGly 24
          |||::::::|||||   |||
      136 TCTGACTGTCCTGCTGGCTGGA 157

  seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-143-470-3

  seq_documentation_block:
  ; Sequence 3, Application US/09143470
  ; Patent No. 6043086
  ; GENERAL INFORMATION:
  ;  APPLICANT: Pan, Yang
  ;  TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR
  ;  FILE REFERENCE: 09404/049001
  ;  CURRENT APPLICATION NUMBER: US/09/143,470
  ;  CURRENT FILING DATE: 1998-08-28
  ;  NUMBER OF SEQ ID NOS: 20
  ;  SOFTWARE: FastSEQ for Windows Version 3.0
  ; SEQ ID NO 3
  ;   LENGTH: 1654
  ;   TYPE: DNA
  ;   ORGANISM: Homo sapiens
  ;   FEATURE:
  ;   NAME/KEY: CDS
  ;   LOCATION: (86)...(1276)
  US-09-143-470-3


  alignment_scores:
              Quality:   48.00            Length:     24
                Ratio:  2.824              Gaps:      0
   Percent Similarity:  70.833   Percent Identity: 45.833
```

```
alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-143-470-3   ..

 Align seg 1/1  to: US-09-143-470-3  from: 1  to: 1654

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         |||         :::|||::::::|||||||||:::      ||||||
      86 ATGGCTCCGATATCTCTGTCGTGGCTGCTCCGCTTGGCCACCTTCTGCCA 135

      17 rLeuAsnThrLeuLeuLeuGly 24
          |||::::::|||||   |||
     136 TCTGACTGTCCTGCTGGCTGGA 157

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-396-479B-7

seq_documentation_block:
; Sequence 7, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2337
US-08-396-479B-7


alignment_scores:
            Quality:   48.00             Length:     15
              Ratio:  3.429               Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 60.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-396-479B-7/rev   ..

 Align seg 1/1  to reverse of: US-08-396-479B-7  from: 1  to: 2406

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeu 22
         |||::::::   ||||||:::|||||||||||||||:::::|||
    2401 TTTTTTTTTTTTTTTTTTTTTTTTGTTCGTTAAATAACTTTTTA 2357

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-818-823-7

seq_documentation_block:
; Sequence 7, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2337
US-08-818-823-7


alignment_scores:
            Quality:   48.00             Length:     15
              Ratio:  3.429               Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 60.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-818-823-7/rev   ..

 Align seg 1/1  to reverse of: US-08-818-823-7  from: 1  to: 2406

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeu 22
         |||::::::   ||||||:::|||||||||||||||:::::|||
    2401 TTTTTTTTTTTTTTTTTTTTTTTTGTTCGTTAAATAACTTTTTA 2357

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-219-932-5

seq_documentation_block:
; Sequence 5, Application US/09219932B
; Patent No. 6153199
; GENERAL INFORMATION:
;  APPLICANT: AUDONNET, Jean-Christophe
;  APPLICANT: BUBLOT, Michel
;  APPLICANT: RIVIERE, Michel
```

```
;  TITLE OF INVENTION: AVIAN RECOMBINANT VACCINE USING, AS VECTOR, THE AVIAN
;  TITLE OF INVENTION: INFECTIOUS LARYNGOTRACHEITIS VIRUS
;  FILE REFERENCE: 454313-2210
;  CURRENT APPLICATION NUMBER: US/09/219,932B
;  CURRENT FILING DATE: 1998-12-23
;  EARLIER APPLICATION NUMBER: PCT/FR97/01138
;  EARLIER FILING DATE: 1997-06-25
;  EARLIER APPLICATION NUMBER: 96/08243
;  EARLIER FILING DATE: 1996-06-27
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3841
;   TYPE: DNA
;   ORGANISM: Infectious Laryngotracheitis Virus
US-09-219-932-5


alignment_scores:
            Quality:   48.00              Length:     14
              Ratio:  4.364                Gaps:      0
 Percent Similarity:  78.571   Percent Identity: 71.429

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-219-932-5   ..

 Align seg 1/1  to: US-09-219-932-5  from: 1  to: 3841

       6 LeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAsn 19
         ||||||||||   ||||||      |||||||||||||:::|||
    2115 TTGGGATTTCAACTCCGACCGGCCATTTTCTGCAGCATAAAC 2156

 seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-623-906A-3

 seq_documentation_block:
 ; Sequence 3, Application US/08623906A
 ; Patent No. 5874217
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Stevenson, Tamara
 ;    APPLICANT:  Dvorak, Jan
 ;    APPLICANT:  Halverson, Joy
 ;    TITLE OF INVENTION:  Microsatellite Sequences for Canine
 ;    TITLE OF INVENTION:  Genotyping
 ;    NUMBER OF SEQUENCES:  60
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 ;      STREET:  4 Embarcadero Center, Suite 3400
 ;      CITY:  San Francisco
 ;      STATE:  CA
 ;      COUNTRY:  US
 ;      ZIP:  94111-4187
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/623,906A
 ;      FILING DATE:
 ;      CLASSIFICATION:  435
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Sherwood, Pamela J.
 ;      REGISTRATION NUMBER:  36,677
 ;      REFERENCE/DOCKET NUMBER:  A-62282/BIR
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  415-781-1989
 ;      TELEFAX:  415-398-3249
 ;  INFORMATION FOR SEQ ID NO:  3:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  279 base pairs
 ;      TYPE:  nucleic acid
 ;      STRANDEDNESS:  double
 ;      TOPOLOGY:  linear
 ;    MOLECULE TYPE:  DNA (genomic)
 ;    FEATURE:
 ;      NAME/KEY:  misc_feature
 ;      LOCATION:  1..111
 ;      OTHER INFORMATION:  /note= "Nucleotides are unique
 ;      OTHER INFORMATION:  flanking sequence"
 ;    FEATURE:
 ;      NAME/KEY:  misc_feature
 ;      LOCATION:  112..204
 ;      OTHER INFORMATION:  /note= "Nucleotides 112-204 are
 ;      OTHER INFORMATION:  repeat sequence"
 ;    FEATURE:
 ;      NAME/KEY:  misc_feature
 ;      LOCATION:  205..279
 ;      OTHER INFORMATION:  /note= "Nucleotides are unique
 ;      OTHER INFORMATION:  flanking sequence"
 US-08-623-906A-3


 alignment_scores:
             Quality:   47.00              Length:     18
               Ratio:  3.133                Gaps:      0
  Percent Similarity:  83.333   Percent Identity: 55.556

 alignment_block:
  US-09-215-435-177_COPY_1_24 x US-08-623-906A-3/rev   ..

  Align seg 1/1  to reverse of: US-08-623-906A-3  from: 1  to: 279

        6 LeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLe 22
          |||:::|||:::|||   ||||||:::|||   ||||||:::   :::||
      174 CTTTCTTTCTTTCTTTCTTTCTTTTTCTTTCTCTTTCCTTCTTTCT 125

       22 uLeu 23
          ||||
      124 TCTT 121

  seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-949-246-3

  seq_documentation_block:
  ; Sequence 3, Application US/08949246
  ; Patent No. 6204035
  ;  GENERAL INFORMATION:
  ;    APPLICANT:  Wiedmer, Therese
  ;    APPLICANT:  Sims, Peter J.
  ;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS TO ALTER THE
  ;    TITLE OF INVENTION:  CELL SURFACE EXPRESSION OF PHOSPHATIDYLSERINE AND OTHER
  ;    TITLE OF INVENTION:  COAGULANT PLASMA MEMBRANE PHOSPHOLIPIDS
  ;    NUMBER OF SEQUENCES:  9
  ;    CORRESPONDENCE ADDRESS:
  ;      ADDRESSEE:  Quarles & Brady
  ;      STREET:  411 East Wisconsin Avenue
  ;      CITY:  Milwaukee
  ;      STATE:  Wisconsin
  ;      COUNTRY:  U.S.A.
  ;      ZIP:  53202-4497
  ;    COMPUTER READABLE FORM:
  ;      MEDIUM TYPE:  Floppy disk
  ;      COMPUTER:  IBM PC compatible
  ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
  ;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
  ;    CURRENT APPLICATION DATA:
  ;      APPLICATION NUMBER:  US/08/949,246
  ;      FILING DATE:
  ;      CLASSIFICATION:  435
  ;    ATTORNEY/AGENT INFORMATION:
  ;      NAME:  Baker, Jean C.
  ;      REGISTRATION NUMBER:  35,433
  ;      REFERENCE/DOCKET NUMBER:  390274.90021
  ;    TELECOMMUNICATION INFORMATION:
  ;      TELEPHONE:  (414) 277-5709
  ;      TELEFAX:  (414) 271-3552
  ;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1622 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-949-246-3


alignment_scores:
            Quality:   47.00            Length:     12
              Ratio:  4.273              Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-949-246-3   ..

 Align seg 1/1  to: US-08-949-246-3  from: 1  to: 1622

      12 PhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         ||||||||||||||||||:::   |||||||||:::
    1456 TTTTTCATCTTTTGTTCAATTTAAACCTTGTTAGTG 1491

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-635-561A-7

seq_documentation_block:
; Sequence 7, Application US/07635561A
; Patent No. 5244805
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Lois K.
;    TITLE OF INVENTION:  Improved Baculovirus Expression Vectors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/635,561A
;      FILING DATE:  19910117
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/353,847
;      FILING DATE:  17-MAY-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 90/02814
;      FILING DATE:  17-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  4-90A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303/499-8080
;      TELEFAX:  303/499-8089
;      TELEX:  823189
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  162 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  92..160
;      OTHER INFORMATION:  /note= "mcs #2 - Multicloning site
;      OTHER INFORMATION:  #2"
US-07-635-561A-7


alignment_scores:
            Quality:   46.50            Length:     27
              Ratio:  2.583              Gaps:      1
 Percent Similarity:  66.667   Percent Identity: 40.741

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-07-635-561A-7   ..

 Align seg 1/1  to: US-07-635-561A-7  from: 1  to: 162

       1 MetLysSerAlaLysLeuGly.............PheLeuLeuArgPh 12
         |||:::   ||||||||||||             |||      :::
       2 ATGTCCCGGGCCAAGCTTGGCGTTATTGAATAAGAATTTAAAAATCAATC 51

      12 ePheIlePheCysSerLeuAsnThrLeuLeu 22
          |||:::::|||:::|||:::|||:::|||
      52 ATTTGTATACTGTAAATTACATACTGTTTTA 82

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-248-335-57

seq_documentation_block:
; Sequence 57, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 57
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-57


alignment_scores:
            Quality:   46.00            Length:     17
              Ratio:  3.067              Gaps:      0
 Percent Similarity:  88.235   Percent Identity: 41.176

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-248-335-57/rev   ..

 Align seg 1/1  to reverse of: US-09-248-335-57  from: 1  to: 960

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGl 24
         |||::::::   ||||||:::|||||||||:::::::::::::   :::||
     913 TTTTTTTTTTTTTTTTTTTTTTTTGTAGTATGAGATCGATTACGATTGG 864

      24 y 24
         |
     863 A 863

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-329-796-2

seq_documentation_block:
; Sequence 2, Application US/09329796
; Patent No. 6291174
; GENERAL INFORMATION:
;  APPLICANT:  Ning LI
;  APPLICANT:  Changxin WU
;  APPLICANT:  Yaofeng ZHAO
;  TITLE OF INVENTION:  DNA MARKERS FOR PIG LITTER SIZE
```

```
;  FILE REFERENCE:  064727.0104
;  CURRENT APPLICATION NUMBER: US/09/329,796
;  CURRENT FILING DATE:  1999-06-10
;  PRIOR APPLICATION NUMBER:  60/088,963
;  PRIOR FILING DATE:  1998-06-10
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1181
;   TYPE: DNA
;   ORGANISM: Sus scrofa
US-09-329-796-2


alignment_scores:
            Quality:   46.00            Length:     21
              Ratio:  2.706              Gaps:      0
 Percent Similarity:  80.952   Percent Identity: 47.619

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-329-796-2/rev   ..

 Align seg 1/1  to reverse of: US-09-329-796-2  from: 1  to: 1181

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||:::      |||   :::|||||||:::|||||||:::|||||||:::
     479 AAAACGTTTCTTCTTATTCTCCTCCTCAATTTTTTTTTTTTTTTGTCAAAG 430

      18 uAsnThrLeuLeu 22
         :::|||:::|||
     429 TAGAACTATATTA 417

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-415-593-44

seq_documentation_block:
; Sequence 44, Application US/08415593
; Patent No. 5912140
; Patent No. 5912140 5776726
;  GENERAL INFORMATION:
;    APPLICANT:  Whoriskey, Susan K.
;    APPLICANT:  Quinn, Cheryl L.
;    APPLICANT:  Tao, Niajun
;    APPLICANT:  Politis-Virk, Karen I.
;    APPLICANT:  Schimmel, Paul R.
;    TITLE OF INVENTION:  Recombinant Pneumoycstis Aminoacyl tRNA
;    TITLE OF INVENTION:  Synthetase Genes, Tester Strains and Assays
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,593
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  CPI94-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-862-9540
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1352 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..1210
US-08-415-593-44


alignment_scores:
            Quality:   46.00            Length:     13
              Ratio:  3.833              Gaps:      0
 Percent Similarity:  92.308   Percent Identity: 53.846

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-415-593-44/rev   ..

 Align seg 1/1  to reverse of: US-08-415-593-44  from: 1  to: 1352

       4 AlaLysLeuGlyPheLeuLeuArgPhePheIlePheCys 16
         :::|||:::   ::::::|||:::|||||||||||||||
    1144 AGCAAGATGAAGTATTTCTTGAAATTCTTTATTTTCTGC 1106

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-123-343A-6

seq_documentation_block:
; Sequence 6, Application US/08123343A
; Patent No. 5593879
;  GENERAL INFORMATION:
;    APPLICANT:  Steller, Hermann
;    APPLICANT:  Abrams, John M.
;    APPLICANT:  Grether, Megan E.
;    APPLICANT:  White, Kristin
;    TITLE OF INVENTION:  Cell Death Genes of Drosophila
;    TITLE OF INVENTION:  Melanogaster and Vertebrate Analogs
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,343A
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/004,957
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-5907A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  61861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-123-343A-6
```

```
alignment_scores:
            Quality:   46.00              Length:     12
              Ratio:  5.111                Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-123-343A-6/rev   ..

 Align seg 1/1  to reverse of: US-08-123-343A-6  from: 1  to: 3900

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsn 19
         ||||||||||||||||||   ||||||      |||
    2354 TTTTTGTTGCGTTTTTTCTGCTTTTGTTTCAACAAT 2319

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-257-073-6

seq_documentation_block:
; Sequence 6, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  966 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-6


alignment_scores:
            Quality:   45.50              Length:     15
              Ratio:  3.792                Gaps:      1
 Percent Similarity:  80.000   Percent Identity: 73.333

alignment_block:
```

```
US-09-215-435-177_COPY_1_24 x US-08-257-073-6/rev   ..

Align seg 1/1  to reverse of: US-08-257-073-6  from: 1  to: 966

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeu 22
         ||||||||||||||||||   ||||||   |||:::   ||||||
     559 TTCTTGCTTCGATTCTTT...TTCTGTTTTCTTCATCTTCTGCTT 518
```

```
OM of: US-09-215-435-177_COPY_1_24 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:54 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence            Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148 +  120.00  301.60  1.9e-07   271
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799 +  120.00  296.13  3.8e-07   499
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799 +  120.00  296.13  3.8e-07    49
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76 +  120.00  295.66  4.0e-07   526
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303 +  120.00  295.44  4.1e-07   539
/cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136 +  120.00  295.44  4.1e-07   539
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-171-625-457 +   60.00  109.79  9.0e+03   264
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-173-464-18981 +   60.00  109.79  9.0e+03     2
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-23127 +   60.00  109.79  9.0e+03     2
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-681-18240 +   60.00  109.79  9.0e+03     2
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-237A-2004 -   60.00   92.74  8.0e+04     1
/cgn1_7/ptodata/1/pna/US096A_COMB.seq:US-09-619-643-31295 +   56.00  145.68  90.25       1
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-350711 +   56.00  145.68  90.25
/cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-350711 +   56.00  145.68  90.25
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-146-905-1329 +   56.00  145.68  90.25      15
/cgn1_7/ptodata/1/pna/US6020_COMB.seq:US-60-209-830-49802 +   56.00  134.11  397.97      5
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-229-511-170 -   56.00   97.85  4.2e+04   327
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-229-511-171 -   56.00   97.84  4.2e+04   327
/cgn1_7/ptodata/1/pna/US6023_COMB.seq:US-60-233-939-74 -   56.00   97.84  4.2e+04  3276
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-226-176-2299 +   56.00   87.33  1.6e+05    10
/cgn1_7/ptodata/1/pna/US6023_COMB.seq:US-60-233-468-2299 +   56.00   87.33  1.6e+05    10
/cgn1_7/ptodata/1/pna/US6031_COMB.seq:US-60-313-371-2299 +   56.00   87.33  1.6e+05    10
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-498-61 -   56.00   80.66  3.8e+05  2225
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-542-62 -   56.00   80.66  3.8e+05  2225
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-245-227-9 -   56.00   80.54  3.8e+05  22553
/cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-112-430-837 -   55.00  138.46  227.74    267
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-478-821-837 -   55.00  138.46  227.74    267
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-229-188417 -   55.00  138.46  227.74
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-866-555-4358 +   55.00  134.42  382.42    419
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-371-805-2173 -   55.00  131.75  538.26    564
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-371-805-2173 -   55.00  131.75  538.26    564
/cgn1_7/ptodata/1/pna/US096B_COMB.seq:US-09-620-608-1790 -   55.00  123.26  1.6e+03    14
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-144-883-1790 -   55.00  123.26  1.6e+03    14
/cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-038-697-568 -   55.00  115.70  4.2e+03   337
/cgn1_7/ptodata/1/pna/US6004_COMB.seq:US-60-046-714-561 -   55.00  115.70  4.2e+03   337
/cgn1_7/ptodata/1/pna/US096B_COMB.seq:US-09-620-392-17378 -   55.00   95.66  5.5e+04     3
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-702-134-25324 -   55.00   95.66  5.5e+04     3
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-213-168-40 +   55.00   95.32  5.8e+04  3276
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-229-512-80 +   55.00   95.32  5.8e+04  3276
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-229-512-125 +   55.00   95.32  5.8e+04   327
/cgn1_7/ptodata/1/pna/US6023_COMB.seq:US-60-234-446-588 -   55.00   95.32  5.8e+04
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-237A-516 -   55.00   92.73  8.0e+04
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-859-775 -   55.00   87.92  1.5e+05
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-803-736-775 -   55.00   87.92  1.5e+05   7
/cgn1_7/ptodata/1/pna/US6025_COMB.seq:US-60-257-537-284 +   55.00   79.56  4.3e+05

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148

seq_documentation_block:
; Sequence 148, Application US/08905279
;  GENERAL INFORMATION:
;    APPLICANT:  Genset SA
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,279
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.011A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  148:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
US-08-905-279-148


alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-08-905-279-148   ..

 Align seg 1/1  to: US-08-905-279-148  from: 1  to: 271

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
```

```
     82 ATTGAATACCCTGTTATTGGGT 103

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799

seq_documentation_block:
; Sequence 5799, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 5799
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 22..318
;   NAME/KEY: sig_peptide
;   LOCATION: 22..93
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
US-09-834-366-5799


alignment_scores:
            Quality:  120.00              Length:      24
              Ratio:  5.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-834-366-5799    ..

 Align seg 1/1  to: US-09-834-366-5799  from: 1  to: 499

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      72 ATTGAATACCCTGTTATTGGGT 93

 seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799

 seq_documentation_block:
 ; Sequence 5799, Application US/60197873
 ; GENERAL INFORMATION:
 ;  APPLICANT:  Bejanin, Stephane
 ;  APPLICANT:  Tanaka, Hiroaki
 ;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Jobert, Severin
 ;  APPLICANT:  Giordano, Jean-Yves
 ;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
 ;  FILE REFERENCE:  81.US1.PRO
 ;  CURRENT APPLICATION NUMBER: US/60/197,873
 ;  CURRENT FILING DATE:  2000-04-18
 ;  NUMBER OF SEQ ID NOS:  52153
 ;  SOFTWARE:  Patent.pm
 ; SEQ ID NO 5799
 ;   LENGTH: 499
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 22..318
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 22..93
 ;   OTHER INFORMATION: Von Heijne matrix
 ;   OTHER INFORMATION: score 4.59999990463257
 ;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
 US-60-197-873-5799


 alignment_scores:
             Quality:  120.00              Length:      24
               Ratio:  5.000                Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-177_COPY_1_24 x US-60-197-873-5799    ..

  Align seg 1/1  to: US-60-197-873-5799  from: 1  to: 499

        1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

       17 rLeuAsnThrLeuLeuLeuGly 24
          ||||||||||||||||||||||
       72 ATTGAATACCCTGTTATTGGGT 93

  seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76

  seq_documentation_block:
  ; Sequence 76, Application US/09215435A
  ; GENERAL INFORMATION:
  ;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
  ;  APPLICANT: Duclert Aymeric,
  ;  APPLICANT: Bougueleret Lydie
  ;  TITLE OF INVENTION: Extended cDNAs
  ;  FILE REFERENCE: GENSET.019A
  ;  CURRENT APPLICATION NUMBER: US/09/215,435A
  ;  CURRENT FILING DATE: 1998-12-17
  ;  EARLIER APPLICATION NUMBER: 60/069,957
  ;  EARLIER FILING DATE: 1997-12-17
  ;  EARLIER APPLICATION NUMBER: 60/074,121
  ;  EARLIER FILING DATE: 1998-2-9
  ;  EARLIER APPLICATION NUMBER: 60/081,563
  ;  EARLIER FILING DATE: 1998-4-13
  ;  EARLIER APPLICATION NUMBER: 60/096,116
  ;  EARLIER FILING DATE: 1998-8-10
  ;  NUMBER OF SEQ ID NOS: 519
  ;  SOFTWARE: Patent.pm
  ; SEQ ID NO 76
  ;   LENGTH: 526
  ;   TYPE: DNA
  ;   ORGANISM: Homo sapiens
  ;   FEATURE:
  ;   NAME/KEY: CDS
  ;   LOCATION: 22..318
  ;   FEATURE:
  ;   NAME/KEY: sig_peptide
  ;   LOCATION: 22..93
  ;   OTHER INFORMATION: Von Heijne matrix
  ;   OTHER INFORMATION: score 4.6
  ;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
  ;   FEATURE:
  ;   NAME/KEY: polyA_signal
  ;   LOCATION: 497..502
  ;   FEATURE:
  ;   NAME/KEY: polyA_site
  ;   LOCATION: 516..526
  US-09-215-435-76
```

```
alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-215-435-76   ..

 Align seg 1/1  to: US-09-215-435-76  from: 1  to: 526

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 71

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      72 ATTGAATACCCTGTTATTGGGT 93

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303

seq_documentation_block:
; Sequence 303, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 303
;   LENGTH: 539
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 32..328
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 32..103
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 508..513
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 528..539
US-09-215-435-303


alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-215-435-303   ..

 Align seg 1/1  to: US-09-215-435-303  from: 1  to: 539

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGT 103

seq_name: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136

seq_documentation_block:
; Sequence 136, Application US/60069957
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  EXTENDED cDNAS FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  381
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/069,957
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.019PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  136:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  32..328
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  508..513
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  528..539
;      IDENTIFICATION METHOD:  blastn
US-60-069-957-136


alignment_scores:
            Quality:  120.00              Length:     24
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-215-435-177_COPY_1_24 x US-60-069-957-136    ..

 Align seg 1/1  to: US-60-069-957-136  from: 1  to: 539

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      32 ATGAAGTCTGCCAAGCTGGGATTTCTTCTAAGATTCTTCATCTTCTGCTC 81

      17 rLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||||||||||||
      82 ATTGAATACCCTGTTATTGGGT 103

seq_name: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-171-625-457

seq_documentation_block:
; Sequence 457, Application US/60171625
; GENERAL INFORMATION:
;  APPLICANT: Skupski, Marian
;  TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS THAT SHOW
;  TITLE OF INVENTION: HIGH HOMOLOGY TO KNOWN HUMAN DISEASE  PROTEINS, AND USES
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: CL000179
;  CURRENT APPLICATION NUMBER: US/60/171,625
;  CURRENT FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 579
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 457
;   LENGTH: 26499
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-60-171-625-457


 alignment_scores:
             Quality:   60.00              Length:      20
               Ratio:  3.750                Gaps:       0
 Percent Similarity:  80.000   Percent Identity: 55.000

 alignment_block:
  US-09-215-435-177_COPY_1_24 x US-60-171-625-457    ..

  Align seg 1/1  to: US-60-171-625-457  from: 1  to: 26499

        2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
          |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
    23740 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 23789

       18 uAsnThrLeu 21
          ||||:::|||
    23790 AAATAGCCTA 23799

  seq_name: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-173-464-18981

  seq_documentation_block:
  ; Sequence 18981, Application US/60173464
  ; GENERAL INFORMATION:
  ;  APPLICANT: Li, Peter W.D.
  ;  TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
  ;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
  ;  TITLE OF INVENTION: THEREOF
  ;  FILE REFERENCE: CL000173
  ;  CURRENT APPLICATION NUMBER: US/60/173,464
  ;  CURRENT FILING DATE: 1999-12-29
  ;  NUMBER OF SEQ ID NOS: 30269
  ;  SOFTWARE: FastSEQ for Windows Version 4.0
  ; SEQ ID NO 18981
  ;   LENGTH: 26499
  ;   TYPE: DNA
  ;   ORGANISM: Drosophila
  US-60-173-464-18981


alignment_scores:
             Quality:   60.00              Length:      20
               Ratio:  3.750                Gaps:       0
 Percent Similarity:  80.000   Percent Identity: 55.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-60-173-464-18981    ..

 Align seg 1/1  to: US-60-173-464-18981  from: 1  to: 26499

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
   23740 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 23789

      18 uAsnThrLeu 21
         ||||:::|||
   23790 AAATAGCCTA 23799

 seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-23127

 seq_documentation_block:
 ; Sequence 23127, Application US/60191637
 ; GENERAL INFORMATION:
 ;  APPLICANT: Venter, J. Craig
 ;  TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
 ;  TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
 ;  TITLE OF INVENTION: USES THEREOF
 ;  FILE REFERENCE: CL000392
 ;  CURRENT APPLICATION NUMBER: US/60/191,637
 ;  CURRENT FILING DATE: 2000-03-23
 ;  NUMBER OF SEQ ID NOS: 42660
 ;  SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 23127
 ;   LENGTH: 26500
 ;   TYPE: DNA
 ;   ORGANISM: DROSOPHILA
 US-60-191-637-23127


 alignment_scores:
             Quality:   60.00              Length:      20
               Ratio:  3.750                Gaps:       0
 Percent Similarity:  80.000   Percent Identity: 55.000

 alignment_block:
  US-09-215-435-177_COPY_1_24 x US-60-191-637-23127    ..

  Align seg 1/1  to: US-60-191-637-23127  from: 1  to: 26500

        2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
          |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
    23741 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 23790

       18 uAsnThrLeu 21
          ||||:::|||
    23791 AAATAGCCTA 23800

  seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-681-18240

  seq_documentation_block:
  ; Sequence 18240, Application US/60191681
  ; GENERAL INFORMATION:
  ;  APPLICANT:  Li, Peter, W.D.
  ;  TITLE OF INVENTION:  ISOLATED DROSOPHILA PROTEINS, NUCLEIC
  ;  TITLE OF INVENTION:  ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
  ;  TITLE OF INVENTION:  USES THEREOF.
  ;  FILE REFERENCE:  cl000390
  ;  CURRENT APPLICATION NUMBER: US/60/191,681
  ;  CURRENT FILING DATE:  2000-03-23
  ;  NUMBER OF SEQ ID NOS:  30973
  ;  SOFTWARE:  FastSEQ for Windows Version 4.0
  ; SEQ ID NO 18240
  ;   LENGTH: 26500
```

```
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-60-191-681-18240


alignment_scores:
            Quality:   60.00              Length:     20
              Ratio:  3.750                Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 55.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-60-191-681-18240   ..

 Align seg 1/1  to: US-60-191-681-18240  from: 1  to: 26500

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
   23741 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 23790

      18 uAsnThrLeu 21
         ||||:::|||
   23791 AAATAGCCTA 23800

seq_name: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-237A-2004

seq_documentation_block:
; Sequence 2004, Application US/09528237A
; GENERAL INFORMATION:
;  APPLICANT:  Venter, J. Craig
;  TITLE OF INVENTION:  The Drosophila Genome, Primary Nucleic
;  TITLE OF INVENTION:  Acid Sequences, Systems Containing The Nucleic Acid
;  TITLE OF INVENTION:  Sequences and Uses Thereof
;  FILE REFERENCE:  CL000284
;  CURRENT APPLICATION NUMBER: US/09/528,237A
;  CURRENT FILING DATE:  2000-03-17
;  NUMBER OF SEQ ID NOS:  2926
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2004
;   LENGTH: 177295
;   TYPE: DNA
;   ORGANISM: Drosophila
US-09-528-237A-2004


alignment_scores:
            Quality:   60.00              Length:     20
              Ratio:  3.750                Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 55.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-528-237A-2004/rev   ..

 Align seg 1/1  to reverse of: US-09-528-237A-2004  from: 1  to: 177295

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         |||   ::::::||||||   ||||||   |||:::|||:::|||   ||
   66807 AAATGTGTTCGTTTAGGCCATCTGCTTTCTTTCTACATTTATTGTTTCCT 66758

      18 uAsnThrLeu 21
         ||||:::|||
   66757 AAATAGCCTA 66748

seq_name: /cgn1_7/ptodata/1/pna/US096A_COMB.seq:US-09-619-643-31295

seq_documentation_block:
; Sequence 31295, Application US/09619643
; GENERAL INFORMATION:
;  APPLICANT:       Fisher, Dane K.
;  APPLICANT:  Lalgudi, Raghunath V.
;  TITLE OF INVENTION:       Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE:       38-21(51230)B
;  CURRENT APPLICATION NUMBER: US/09/619,643
;  CURRENT FILING DATE:  2000-07-19
;  NUMBER OF SEQ ID NOS:       32236
; SEQ ID NO 31295
;   LENGTH: 158
;   TYPE: DNA
;   ORGANISM: Zea mays
;   OTHER INFORMATION: Clone ID: LIB3181-011-P1-K2-H4
US-09-619-643-31295


alignment_scores:
            Quality:   56.00              Length:     16
              Ratio:  4.308                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-619-643-31295   ..

 Align seg 1/1  to: US-09-619-643-31295  from: 1  to: 158

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         |||||||||   ||||||||||||||||||||      |||:::|||
     100 TTCTTGTTAAGTTTTTTTATTTTCTGTTCCTTGTTTTGTCTTTTTCTT 147

seq_name: /cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-350711

seq_documentation_block:
; Sequence 350711, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 350711
;   LENGTH: 158
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-654-617-350711


alignment_scores:
            Quality:   56.00              Length:     16
              Ratio:  4.308                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-654-617-350711   ..

 Align seg 1/1  to: US-09-654-617-350711  from: 1  to: 158

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         |||||||||   ||||||||||||||||||||      |||:::|||
     100 TTCTTGTTAAGTTTTTTTATTTTCTGTTCCTTGTTTTGTCTTTTTCTT 147

seq_name: /cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-350711

seq_documentation_block:
; Sequence 350711, Application US/09684016
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/684,016
;  CURRENT FILING DATE:  2000-10-10
;  PRIOR APPLICATION NUMBER:       US 09/654,617
;  PRIOR FILING DATE:       2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 350711
;   LENGTH: 158
```

```
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-684-016-350711


alignment_scores:
            Quality:   56.00              Length:     16
              Ratio:  4.308                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-684-016-350711   ..

 Align seg 1/1  to: US-09-684-016-350711  from: 1  to: 158

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         |||||||||   ||||||||||||||||||||      |||:::|||
     100 TTCTTGTTAAGTTTTTTTATTTTCTGTTCCTTGTTTTGTCTTTTTCTT 147

seq_name: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-146-905-1329

seq_documentation_block:
; Sequence 1329, Application US/60146905
; GENERAL INFORMATION:
;  APPLICANT: Cheikh, Nordine
;  APPLICANT: Deng, Molian
;  APPLICANT: Fisher, Dane K.
;  APPLICANT: Miller, Philip W.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION: PLANTS
;  FILE REFERENCE: 38-21(51232)A
;  CURRENT APPLICATION NUMBER: US/60/146,905
;  CURRENT FILING DATE: 1999-08-02
;  NUMBER OF SEQ ID NOS: 2850
; SEQ ID NO 1329
;   LENGTH: 158
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3181-011-P1-K2-H4
US-60-146-905-1329


alignment_scores:
            Quality:   56.00              Length:     16
              Ratio:  4.308                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-60-146-905-1329   ..

 Align seg 1/1  to: US-60-146-905-1329  from: 1  to: 158

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         |||||||||   ||||||||||||||||||||      |||:::|||
     100 TTCTTGTTAAGTTTTTTTATTTTCTGTTCCTTGTTTTGTCTTTTTCTT 147
```

```
OM of: US-09-215-435-177_COPY_1_24 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:08 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000


score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-708-319 +   53.00  128.73  5.90      4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8488 +   51.00  128.46  6.11
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-9659 +   51.00  126.75  7.61
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-11087 +   51.00  126.21  8.15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-22094 +   51.00  125.45  8.99
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-22403 +   51.00  124.94  9.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-18975 -   51.00  124.88  9.67
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-9295 +   50.00  112.73  45.95
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-9530 +   50.00  112.73  45.95
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-214 +   50.00  104.43  133.23     3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8149 +   49.00  125.81  8.58
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-10026 +   49.00  121.22  15.46
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-5952 +   48.00  127.75  6.69
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-4662 -   47.00  114.28  37.66
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-111 +   47.00  111.88  51.26      5
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1444 +   47.00  111.88  51.26
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1030 -   46.00  115.10  33.91
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1758 -   46.00  115.10  33.91
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-760 -   46.00  105.82  111.49     8
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-875 +   46.00  104.78  127.31     9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-694 -   46.00  101.75  187.82     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-937-974-25 -   46.00   97.46  325.65    22
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-795-668-667 -   45.00  109.93  65.78      4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-795-668-668 -   45.00  109.93  65.78      4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5520 -   45.00  109.44  70.04
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-103 +   45.00  102.27  175.72     9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-522-2 +   45.00   88.35  1.0e+03   468
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-347-172 -   45.00   85.39  1.5e+03     6
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-95 -   45.00   85.38  1.5e+03    65
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-795-668-1 -   45.00   37.71  2.4e+05   150
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-834-975-773 +   44.50   95.36  426.44     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1830 -   44.50   91.29  718.70
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-353 -   44.50   71.93  8.5e+03     2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2902 -   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2903 -   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2904 -   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2905 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2906 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2907 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2908 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2909 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2910 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2911 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2912 +   44.50   71.93  8.5e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2913 +   44.50   71.93  8.5e+03

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-708-319

seq_documentation_block:
; Sequence 319, Application US/09969708
; GENERAL INFORMATION:
;  APPLICANT: Augustus, Meena
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-70
;  CURRENT APPLICATION NUMBER: US/09/969,708
;  CURRENT FILING DATE:  2001-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,606
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,608
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,425
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 319
;   LENGTH: 473
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-969-708-319


alignment_scores:
            Quality:   53.00             Length:     17
              Ratio:  3.786               Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 52.941

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-969-708-319   ..

 Align seg 1/1  to: US-09-969-708-319  from: 1  to: 473

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGl 24
         |||:::|||   :::|||:::||||||:::   ||||||||||:::   ||
      15 TTTATTTTATTTTATTTTTTTTTTTGTCAGACAAACACATTGATTTCTGG 64

      24 y 24
         |
      65 A 65

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8488

seq_documentation_block:
; Sequence 8488, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8488
;   LENGTH: 274
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-8488


alignment_scores:
            Quality:   51.00             Length:     14
```

```
            Ratio:  4.250                Gaps:      0
Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
US-09-215-435-177_COPY_1_24 x US-09-388-906A-8488    ..

 Align seg 1/1  to: US-09-388-906A-8488  from: 1  to: 274

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   |||||||:::|||||||   |||:::::|||
     173 TTTCTTCTTTCTTTTTTTTTATTTTGTTGTCTGTCTTCTTTG 214

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-9659

seq_documentation_block:
; Sequence 9659, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9659
;   LENGTH: 333
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-9659


alignment_scores:
            Quality:   51.00              Length:      14
              Ratio:  4.250                Gaps:       0
Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
US-09-215-435-177_COPY_1_24 x US-09-388-906A-9659    ..

 Align seg 1/1  to: US-09-388-906A-9659  from: 1  to: 333

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   |||||||:::|||||||   |||:::::|||
     178 TTTCTTCTTTCTTTTTTTTTATTTTGTTGTCTGTCTTCTTTG 219

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-11087

seq_documentation_block:
; Sequence 11087, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11087
;   LENGTH: 354
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-11087


alignment_scores:
            Quality:   51.00              Length:      14
              Ratio:  4.250                Gaps:       0
Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
US-09-215-435-177_COPY_1_24 x US-09-388-906A-11087    ..

 Align seg 1/1  to: US-09-388-906A-11087  from: 1  to: 354

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   |||||||:::|||||||   |||:::::|||
     199 TTTCTTCTTTCTTTTTTTTTATTTTGTTGTCTGTCTTCTTTG 240

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-22094

seq_documentation_block:
; Sequence 22094, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22094
;   LENGTH: 386
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-22094


alignment_scores:
            Quality:   51.00              Length:      14
              Ratio:  4.250                Gaps:       0
Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
US-09-215-435-177_COPY_1_24 x US-09-388-906A-22094    ..

 Align seg 1/1  to: US-09-388-906A-22094  from: 1  to: 386

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   |||||||:::|||||||   |||:::::|||
     241 TTTCTTCTTTCCTTCTTTTTATTTTGTTGTCTGTCTTCTTTG 282

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-22403

seq_documentation_block:
; Sequence 22403, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22403
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-22403


alignment_scores:
            Quality:   51.00              Length:      14
              Ratio:  4.250                Gaps:       0
Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
US-09-215-435-177_COPY_1_24 x US-09-388-906A-22403    ..
```

```
 Align seg 1/1  to: US-09-388-906A-22403  from: 1  to: 409

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   ||||||:::||||||   |||:::::::|||
     280 TTTCTTCTTTCCTTCTTTTATTTTGTTGTCTGTCTTCTTTG 321

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-18975

seq_documentation_block:
; Sequence 18975, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18975
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-18975


alignment_scores:
            Quality:   51.00             Length:     24
              Ratio:  3.188               Gaps:      0
 Percent Similarity:  66.667   Percent Identity: 45.833

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-388-906A-18975/rev   ..

 Align seg 1/1  to reverse of: US-09-388-906A-18975  from: 1  to: 412

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         :::   |||   :::|||:::|||   :::   |||   |||   |||
     408 CTCCACTCTTTGTCTTTGTCGTTTTGTATCACCTTCAGCATTCAATGCCA 359

      17 rLeuAsnThrLeuLeuLeuGly 24
          ||||||:::|||   ||||||
     358 TTTGAATAGCTTGTACTTGGGA 337

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-9295

seq_documentation_block:
; Sequence 9295, Application US/09815242
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9295
;   LENGTH: 1233
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1233)
US-09-815-242-9295


alignment_scores:
            Quality:   50.00             Length:     17
              Ratio:  3.571               Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 47.059

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-815-242-9295   ..

 Align seg 1/1  to: US-09-815-242-9295  from: 1  to: 1233

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         |||||||||::::::   ::::::||||||   ||||||   :::|||::
    1088 GGATTTTTGATGGAAGTGATGGTGTTTTGCGTTTTAAACAGAATTTTAAT 1137

      23 u 23
         :
    1138 G 1138

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-9530

seq_documentation_block:
; Sequence 9530, Application US/09815242
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9530
;   LENGTH: 1233
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1)...(1233)
US-09-815-242-9530


alignment_scores:
            Quality:   50.00              Length:     17
              Ratio:  3.571                Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 47.059

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-815-242-9530   ..

 Align seg 1/1  to: US-09-815-242-9530  from: 1  to: 1233

       7 GlyPheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLe 23
         |||||||||:::::   ::::::||||||   ||||||   :::|||::
    1088 GGATTTTTGATGGAAGTGATGGTGTTTTGCGTTTTAAACAGAATTTTAAT 1137

      23 u 23
         :
    1138 G 1138

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-214

seq_documentation_block:
; Sequence 214, Application US/09971392
; GENERAL INFORMATION:
;  APPLICANT: Peterson, David P.
;  APPLICANT:  Pearson, Cecelia I.
;  APPLICANT:  Cocks, Benjamin G.
;  TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
;  FILE REFERENCE: PA-0029 US
;  CURRENT APPLICATION NUMBER: US/09/971,392
;  CURRENT FILING DATE:  2001-10-03
;  PRIOR APPLICATION NUMBER: 60/237,652
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 260
;  SOFTWARE: PERL Program
; SEQ ID NO 214
;   LENGTH: 3175
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Template ID: 332263.1
;   NAME/KEY: unsure
;   LOCATION: 1854-1892
;   OTHER INFORMATION: a, t, c, g, or other
US-09-971-392-214


alignment_scores:
            Quality:   50.00              Length:     14
              Ratio:  4.167                Gaps:      0
 Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-971-392-214   ..

 Align seg 1/1  to: US-09-971-392-214  from: 1  to: 3175

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||   :::||||||||||||||||||||||   :::::::|||
     939 TTCAGAATCAGATTCTTCATTTTTTGCAGCTACGATGTTCTT 980

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8149

seq_documentation_block:
; Sequence 8149, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8149
;   LENGTH: 208
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-8149


alignment_scores:
            Quality:   49.00              Length:     14
              Ratio:  4.083                Gaps:      0
 Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-388-906A-8149   ..

 Align seg 1/1  to: US-09-388-906A-8149  from: 1  to: 208

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   ||||||:::||||||   |||:::::::|||
     111 TTTCTTCTTTCTTTTTTTTTTTTTGTTGTCTGTCTTCTTTG 152

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-10026

seq_documentation_block:
; Sequence 10026, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10026
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-10026


alignment_scores:
            Quality:   49.00              Length:     14
              Ratio:  4.083                Gaps:      0
 Percent Similarity:  85.714   Percent Identity: 64.286

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-388-906A-10026   ..

 Align seg 1/1  to: US-09-388-906A-10026  from: 1  to: 351

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
         |||||||||   ||||||:::||||||   |||:::::::|||
     209 TTTCTTCTTTCTTTTTTTTTTTTTGTTGTCTGTCTTCTTTG 250

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-5952

seq_documentation_block:
; Sequence 5952, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
```

```
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE:  1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5952
;   LENGTH: 125
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-5952


alignment_scores:
           Quality:   48.00            Length:     12
             Ratio:  4.364              Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-388-906A-5952   ..

 Align seg 1/1  to: US-09-388-906A-5952  from: 1  to: 125

       6 LeuGlyPheLeuLeuArgPhePheIlePheCysSer 17
         |||||||::   |||||||||||||:::|||||||||
      22 CTAGGCATTTCTCTGCGATTCTTCTTCTTCTGCTCG 57

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-4662

seq_documentation_block:
; Sequence 4662, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4662
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(435)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-4662


alignment_scores:
           Quality:   47.00            Length:     23
             Ratio:  2.611              Gaps:      0
 Percent Similarity:  78.261   Percent Identity: 39.130

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-922-340-4662/rev   ..

 Align seg 1/1  to reverse of: US-09-922-340-4662  from: 1  to: 435

       1 MetLysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSe 17
         ::::::|||:::   |||   ||||||         |||||||:::|||::
     234 CTTAGATCAAGTTACTTAACCTTTTTGAGCTTCAGTTTTATCATTTGTAA 185

      17 rLeuAsnThrLeuLeuLeu 23
         :::::::|||:::|||:::
     184 AATGGATACAATCCTGATG 166

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-111

seq_documentation_block:
; Sequence 111, Application US/09954456
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 111
;   LENGTH: 572
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-111


alignment_scores:
           Quality:   47.00            Length:     15
             Ratio:  3.615              Gaps:      0
 Percent Similarity:  86.667   Percent Identity: 60.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x US-09-954-456-111   ..

 Align seg 1/1  to: US-09-954-456-111  from: 1  to: 572

       9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeu 23
         :::||||||:::|||:::|||   |||||||||   :::|||||||
     409 GTATTGCGACTGTTTGTTTTTTCTTCTTTAAATATTATTTTGCTT 453
```



```
OM of: US-09-215-435-177_COPY_1_24 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:30 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_1_24
Query length: 24
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est50:AW697731      +   62.00   150.09   30.55     603  ! AW697731 Str1-A6 Sugar Beet ger
gb_gss3:CNS02GG5       +   62.00   145.94   51.99     980  ! AL196286 Tetraodon nigroviridis
gb_gss32:AZ760839      +   59.00   144.34   63.84     487  ! AZ760839 1M0554B19R Mouse 10kb
gb_gss23:AZ222099      +   59.00   143.95   67.15     510  ! AZ222099 RPCI-23-63O3.TV RPCI-2
gb_est62:BB351590      +   58.50   147.72   41.42     283  ! BB351590 BB351590 RIKEN full-le
gb_gss11:AQ414393      +   58.00   142.23   83.68     464  ! AQ414393 RPCI-11-169M20.TJ RPCI
gb_est70:BE117416      -   57.50   144.93   59.25     292  ! BE117416 UI-R-BS1-ayi-e-09-0-UI
gb_gss11:AQ395192      +   57.50   139.65   116.62    542  ! AQ395192 CITBI-E1-2556P5.TR CIT
gb_gss11:AQ393358      +   57.50   138.94   127.74    589  ! AQ393358 CITBI-E1-2556D1.TR CIT
gb_est24:AI739690      +   56.00   141.59   90.87     277  ! AI739690 ME000111.FOR Egg stage
gb_gss13:AQ562045      +   56.00   138.96   127.35    377  ! AQ562045 HS_5237_A2_D10_SP6E RP
gb_est100:BG399970     -   56.00   131.50   331.40    903  ! BG399970 602442023F1 NIH_MGC_75
gb_gss37:PT010M15R     +   56.00   130.78   363.68    983  ! AL448605 Paramecium tetraurelia
gb_gss18:AQ879006      +   56.00   127.03   587.80    1524  ! AQ879006 HS_3154_A2_E10_T7C CI
gb_est98:BG287285      -   55.50   128.98   458.14    1047  ! BG287285 602382223F1 NIH_MGC_9
gb_est18:AI271509      -   55.00   140.96   98.57     222  ! AI271509 qw83f07.x1 NCI_CGAP_Br
gb_est57:BB142203      -   55.00   139.77   114.72    255  ! BB142203 BB142203 RIKEN full-le
gb_gss21:AZ134350      +   55.00   137.57   152.14    330  ! AZ134350 OSJNBb0112C05f CUGI Ri
gb_est65:BB447417      +   55.00   137.52   153.15    332  ! BB447417 BB447417 RIKEN full-le
gb_gss32:AZ761030      -   55.00   137.32   157.19    340  ! AZ761030 1M0555F12F Mouse 10kb
gb_gss7:AQ087208       +   55.00   137.10   161.75    349  ! AQ087208 HS_2200_A1_B07_MR CIT
gb_est41:AW006889      +   55.00   135.72   192.95    410  ! AW006889 ws15f07.x1 NCI_CGAP_Ki
gb_est2:AA127637       +   55.00   135.53   197.60    419  ! AA127637 zn94c12.r1 Stratagene
gb_est90:BF606460      -   55.00   134.93   213.66    450  ! BF606460 273506 MARC 3BOV Bos t
gb_gss37:TA133G03Q     +   55.00   132.92   276.25    569  ! AL466224 T. brucei sheared genc
gb_htc:AK020315        +   55.00   132.31   298.65    611  ! AK020315 Mus musculus adult mal
gb_gss35:B67216        +   55.00   132.24   301.33    616  ! B67216 CpG0033B CpIOWAgDNA1 Cry
gb_gss26:AZ409586      +   55.00   131.99   311.52    635  ! AZ409586 1M0181G15F Mouse 10kb
gb_est75:BE532096      -   55.00   129.83   410.51    817  ! BE532096 601235518F1 NCI_CGAP_M
gb_est90:BF668274      -   55.00   128.77   470.28    925  ! BF668274 602122471F1 NIH_MGC_56
gb_est58:BB175242      +   54.50   137.10   161.72    301  ! BB175242 BB175242 RIKEN full-le
gb_gss15:AQ715014      -   54.50   131.80   319.15    560  ! AQ715014 HS_5417_B2_A12_T7A RPC
gb_est72:BE225485      -   54.50   130.03   400.47    689  ! BE225485 MD0393 Meloidogyne inc
gb_est58:BB161839      +   54.00   137.35   156.52    252  ! BB161839 BB161839 RIKEN full-le
gb_est8:AA493666       +   54.00   136.07   184.60    293  ! AA493666 nh01e05.s1 NCI_CGAP_Th
gb_est65:BB452697      +   54.00   135.75   192.21    304  ! BB452697 BB452697 RIKEN full-le
gb_est47:AW447086      -   54.00   131.55   329.24    497  ! AW447086 88203 MARC 1BOV Bos ta
gb_est53:AW936552      -   54.00   131.55   329.24    497  ! AW936552 QV4-DT0021-220200-112-
gb_est53:AW908772      +   54.00   131.52   330.69    499  ! AW908772 uf57c08.y1 Soares_mamm
gb_gss14:AQ662206      -   54.00   131.37   337.23    508  ! AQ662206 HS_5418_B2_D08_SP6E RP
gb_gss15:AQ679237      +   54.00   131.07   350.34    526  ! AQ679237 HS_5417_A1_A12_T7A RPC
gb_gss28:AZ512307      -   54.00   130.26   388.43    578  ! AZ512307 1M0357F15R Mouse 10
gb_gss18:AQ931141      -   54.00   129.80   412.04    610  ! AQ931141 RPCI-23-267L24.TJ R
gb_est45:AW293528      -   54.00   129.78   413.52    612  ! AW293528 UI-H-BI2-ahq-c-10-0
gb_gss25:AZ348935      -   54.00   129.60   423.14    625  ! AZ348935 1M0085P12R Mouse 10

seq_name: gb_est50:AW697731

seq_documentation_block:
LOCUS       AW697731      603 bp    mRNA            EST       17-APR-2000
DEFINITION  Str1-A6 Sugar Beet germination cDNA library Beta vulgaris cDNA
            clone #6 5' similar to 60S acidic ribosomal protein (RP1), mRNA
            sequence.
ACCESSION   AW697731
VERSION     AW697731.1  GI:7580174
KEYWORDS    EST.
SOURCE      beet.
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   de los Reyes,B.G., McGrath,J.M., Myers,S. and Derrico,C.
  TITLE     Differential gene expression in sugar beet seedlings (Beta vulgaris
            ) germinated under stress conditions
  JOURNAL   Unpublished (2000)
COMMENT     Contact: J. Mitchell McGrath
            Sugar Beet Genetics Lab., USDA/ARS/Sugar Beet and Bean Research
            Unit
            Michigan State University
            494 Plant and Soil Sci. Bldg., East Lansing MI 48824, USA
            Tel: (517)-432-2355
            Fax: (517)-337-6782
            Email: mitchmcg@pilot.msu.edu
            Seq primer: T3
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Beta vulgaris"
                     /cultivar="USH20"
                     /db_xref="taxon:3555"
                     /clone="#6"
                     /clone_lib="Sugar Beet germination cDNA library"
                     /tissue_type="whole seedlings"
                     /dev_stage="4-day germination under stress (salt/NaCl,
                     dehydration/Mannitol and anaerobic stress)"
                     /lab_host="SOLR"
                     /note="Organ: seeds; Vector: pBluescript II KS (+) excised
                     from Lambda Uni-Zap XR; Site_1: EcoRI; Site_2: XhoI; cDNAs
                     were derived from reverse transcription of pooled mRNA
                     samples  from stressed seedlings (germinated for 4 days in
                     150mM NaCl, 200mM  Mannitol, submerged in distilled water
                     and 0.3% Hydrogen peroxide).  The cDNA library was
                     generated by directional ligation of the cDNAs  in the
                     EcoRI and XhoI sites of Lambda Uni-Zap XR vector
                     (Stratagene).  The library was excised as a pBluescript
                     plasmid from Lambda Uni-Zap  XR by ExAssist helper phage
                     and plated on SOLR host cells. "
BASE COUNT      144 a    121 c    149 g    189 t
ORIGIN


alignment_scores:
           Quality:   62.00              Length:     17
             Ratio:  3.647                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 58.824

alignment_block:
 US-09-215-435-177_COPY_1_24 x AW697731   ..

 Align seg 1/1  to: AW697731  from: 1  to: 603

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGl 24
         ||||||||||:::|||||||:::|||||||:::|||:::::::::::::|||||
```

```
    474 TTTTTGTTAAAGTTTTTTTTTTTTTGTACTCTGAGATCAGTTATTTTGGG 523

     24 y 24
        |
    524 G 524

seq_name: gb_gss3:CNS02GG5

seq_documentation_block:
LOCUS       CNS02GG5      980 bp    DNA             GSS       13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            137H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL196286
VERSION     AL196286.1  GI:7834436
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 980)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 980)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 980)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .980
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="137H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG137CD11LP1-end : T7"
BASE COUNT      285 a    171 c    184 g    335 t      5 others
ORIGIN


alignment_scores:
            Quality:   62.00            Length:     23
              Ratio:  3.647              Gaps:      0
 Percent Similarity:  73.913   Percent Identity: 52.174

alignment_block:
 US-09-215-435-177_COPY_1_24 x CNS02GG5   ..

 Align seg 1/1  to: CNS02GG5  from: 1  to: 980

      2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
        |||     ||||||  |||||||||:::||||||:::||||||   ::
    835 AAATGTCCCAAACTATTTTTTTTATTACAATTTTTTGTTTTTTGCTGTTT 884

     18 uAsnThrLeuLeuLeuGly 24
        ::::     |||:::|||
    885 CTCCATCTGGCTTGTCGGC 903

seq_name: gb_gss32:AZ760839

seq_documentation_block:
LOCUS       AZ760839      487 bp    DNA             GSS       16-FEB-2001
DEFINITION  1M0554B19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0554B19 R, DNA sequence.
ACCESSION   AZ760839
VERSION     AZ760839.1  GI:12869114
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0554  row: B  column: 19
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0554B19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       94 a     80 c     90 g    223 t
ORIGIN


alignment_scores:
            Quality:   59.00            Length:     16
              Ratio:  3.933              Gaps:      0
 Percent Similarity:  93.750   Percent Identity: 62.500

alignment_block:
 US-09-215-435-177_COPY_1_24 x AZ760839   ..
```

```
 Align seg 1/1  to: AZ760839  from: 1  to: 487

      9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeuLeuLeuGly 24
        ||||||   ||||||:::|||||||:::|||||||:::|||:::::|||
    219 CTTTTAATATTCTTTCTTTTCTGTACATTAAATGTTTTGATTATGGGG 266

seq_name: gb_gss23:AZ222099

seq_documentation_block:
LOCUS       AZ222099      510 bp    DNA             GSS       14-JUN-2000
DEFINITION  RPCI-23-63O3.TV RPCI-23 Mus musculus genomic clone RPCI-23-63O3,
            DNA sequence.
ACCESSION   AZ222099
VERSION     AZ222099.1  GI:8530064
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-63O3.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 63  row: O  column: 3
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-63O3"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      114 a    126 c     72 g    198 t
ORIGIN


alignment_scores:
            Quality:   59.00            Length:     19
              Ratio:  3.933              Gaps:      0
 Percent Similarity:  78.947   Percent Identity: 68.421

alignment_block:
 US-09-215-435-177_COPY_1_24 x AZ222099   ..

 Align seg 1/1  to: AZ222099  from: 1  to: 510

      3 SerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAs 19
        |||   ||||||   ||||||:::|||||||||||||   |||   |||||
    177 AGTCTTAAACTTTAATTTCTTTTTAGATTTTTCATAGAATGTTTATTGAA 226

     19 nThrLeu 21
        ||||:::
    227 CACAGTC 233

seq_name: gb_est62:BB351590

seq_documentation_block:
LOCUS       BB351590      283 bp    mRNA            EST       12-JUL-2000
DEFINITION  BB351590 RIKEN full-length enriched, 10 days neonate cerebellum Mus
            musculus cDNA clone B930085F07 3', mRNA sequence.
ACCESSION   BB351590
VERSION     BB351590.1  GI:9063418
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno;Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="B930085F07"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     cerebellum"
                     /sex="mixed"
                     /tissue_type="cerebellum"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
```

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 50 a 71 c 50 g 112 t
ORIGIN

alignment_scores:
Quality: 58.50 Length: 21
Ratio: 3.900 Gaps: 1
Percent Similarity: 71.429 Percent Identity: 61.905

alignment_block:
US-09-215-435-177_COPY_1_24 x BB351590 ..

Align seg 1/1 to: BB351590 from: 1 to: 283

```
      6 LeuGlyPheLeuLeuArgPhePheIlePheCys.........SerLeuAs 19
        |||||||||   ||||||||||||:::|||||||          ||||||||
     14 TTGGGCTTCACCCTTCGGTTTTTTGTTTTTTGTTTTTTCTGGTCCCTGAA 63

     19 nThrLeuLeuLeu 23
        |      :::|||
     64 TCTGGCATTTCTG 76
```

seq_name: gb_gss11:AQ414393

seq_documentation_block:
LOCUS AQ414393 464 bp DNA GSS 23-MAR-1999
DEFINITION RPCI-11-169M20.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-169M20 , DNA sequence.
ACCESSION AQ414393
VERSION AQ414393.1 GI:4470175
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter ,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .464
/organism="Homo sapiens"
/db_xref="GDB:7564819"
/db_xref="taxon:9606"
/clone="RPCI-11-169M20"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 106 a 96 c 69 g 192 t 1 others
ORIGIN

alignment_scores:
Quality: 58.00 Length: 22
Ratio: 3.222 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 54.545

alignment_block:
US-09-215-435-177_COPY_1_24 x AQ414393 ..

Align seg 1/1 to: AQ414393 from: 1 to: 464

```
      3 SerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLeuAs 19
        |||:::   |||   ||||||:::   ||||||:::||||||||||:::
    121 TCCTGTATCCTCTTTTTTCTTTTTTTTTTTTTTTTGTTTTGTTCTGTTCC 170

     19 nThrLeuLeuLeuGly 24
        |||:::::::||||||
    171 AACTTTTATTTTAGGA 186
```

seq_name: gb_est70:BE117416

seq_documentation_block:
LOCUS BE117416 292 bp mRNA EST 13-JUN-2000
DEFINITION UI-R-BS1-ayi-e-09-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone UI-R-BS1-ayi-e-09-0-UI 3', mRNA sequence.
ACCESSION BE117416
VERSION BE117416.1 GI:8509521
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 292)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .292

```
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BS1-ayi-e-09-0-UI"
                     /clone_lib="UI-R-BS1"
                     /dev_stage="embryonic 13 dpc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS1
                     library is derived from 13 dpc whole embryo tissue.  For a
                     detailed description of the library from which this clone
                     was derived, please visit our web site at
                     ratest.eng.uiowa.edu.
                     TAG_LIB=UI-R-BS1
                     TAG_TISSUE=embryo at 13 dpc
                     TAG_SEQ=AATCC"
BASE COUNT       74 a     40 c     83 g     94 t      1 others
ORIGIN


alignment_scores:
           Quality:   57.50             Length:     21
             Ratio:  3.382               Gaps:      1
 Percent Similarity:  80.952   Percent Identity: 61.905

alignment_block:
 US-09-215-435-177_COPY_1_24 x BE117416/rev   ..

 Align seg 1/1  to reverse of: BE117416  from: 1  to: 292

       6 LeuGlyPheLeuLeuArgPhePheIlePheCys.........SerLeuAs 19
         ||||||||||||||||||||||||:::|||             :::|||||
     104 CTTGGATTCTTACTTCGTTTTTTTTTTTTCCTGAAAGCGAAAGCTTTAAA 55

      19 nThrLeuLeuLeu 23
         ||||::::::|||
      54 TACTGTGATTCTA 42

seq_name: gb_gss11:AQ395192

seq_documentation_block:
LOCUS       AQ395192      542 bp    DNA             GSS       06-MAR-1999
DEFINITION  CITBI-E1-2556P5.TR CITBI-E1 Homo sapiens genomic clone 2556P5, DNA
            sequence.
ACCESSION   AQ395192
VERSION     AQ395192.1  GI:4366218
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2556P5"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      142 a     88 c     88 g    224 t
ORIGIN


alignment_scores:
           Quality:   57.50             Length:     25
             Ratio:  3.594               Gaps:      1
 Percent Similarity:  64.000   Percent Identity: 48.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AQ395192   ..

 Align seg 1/1  to: AQ395192  from: 1  to: 542

       7 GlyPheLeu.....................LeuArgPhePheIlePheCy 16
         |||||||||                     |||   :::|||:::|||||
     144 GGTTTTCTGTCTCTGGCCTTCAAGATTTTCCTTTTGTATTTTCTTTTCTG 193

      16 sSerLeuAsnThrLeuLeuLeuGly 24
         ||||||||||||:::   :::|||
     194 TAGTTTAAATACAATATGCGTAGGG 218

seq_name: gb_gss11:AQ393358

seq_documentation_block:
LOCUS       AQ393358      589 bp    DNA             GSS       06-MAR-1999
DEFINITION  CITBI-E1-2556D1.TR CITBI-E1 Homo sapiens genomic clone 2556D1, DNA
            sequence.
ACCESSION   AQ393358
VERSION     AQ393358.1  GI:4364381
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2556D1"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      161 a     93 c     98 g    237 t
ORIGIN
```

```
alignment_scores:
            Quality:   57.50              Length:     25
              Ratio:  3.594                Gaps:      1
 Percent Similarity:  64.000   Percent Identity: 48.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AQ393358   ..

 Align seg 1/1  to: AQ393358  from: 1  to: 589

      7 GlyPheLeu.....................LeuArgPhePheIlePheCy 16
        |||||||||                     |||   :::|||:::|||||
    143 GGCTTTCTGTCTCTGGCCTTCAAGATTTTCCTTTTGTATTTTCTTTTCTG 192

     16 sSerLeuAsnThrLeuLeuLeuGly 24
        |||||||||||||:::   :::|||
    193 TAGTTTAAATACAATATGCGTAGGG 217

seq_name: gb_est24:AI739690

seq_documentation_block:
LOCUS       AI739690      277 bp    mRNA            EST       21-JUN-1999
DEFINITION  ME000111.FOR Egg stage cDNA expression library in Lambda ZAPII
            Schistosoma mansoni cDNA 5', mRNA sequence.
ACCESSION   AI739690
VERSION     AI739690.1  GI:5108021
KEYWORDS    EST.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
            Trematoda; Digenea; Strigeidida; Schistosomatoidea;
            Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   Oliveira,G.C. and Baba,J.
  TITLE     Cataloguing Schistosoma mansoni genes with expressed sequence tags
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Oliveira, Guilherme
            Lab. Parasitologia Cel. e Mol.
            Centro de Pesquisas Rene Rachou - FIOCRUZ
            Av. Augusto de Lima 1715 Barro Preto, Belo Horizonte, MG, CEP 30190
            , Brazil
            Tel: 55 31 2953566
            Fax: 55 31 2952115
            Email: oliveira@netra.cpqrr.fiocruz.br
            Insert Length: 277   Std Error: 0.00
            Seq primer: M13 Reverse Universal Sequencing primer.
FEATURES             Location/Qualifiers
     source          1. .277
                     /organism="Schistosoma mansoni"
                     /strain="LE"
                     /db_xref="taxon:6183"
                     /clone_lib="Egg stage cDNA expression library in Lambda
                     ZAPII"
                     /sex="Mixed"
                     /dev_stage="Egg"
                     /note="Vector: Uni-Zap XR vector, Stratagene (pBluescript
                     SK).; Site_1: EcoRI; Site_2: XhoI; mRNA was extracted from
                     eggs and the library was constructed  and excised
                     according to the manufacturer's instructions."
BASE COUNT       75 a     53 c     47 g    102 t
ORIGIN


alignment_scores:
            Quality:   56.00              Length:     13
              Ratio:  4.667                Gaps:      0
 Percent Similarity:  92.308   Percent Identity: 69.231

alignment_block:
 US-09-215-435-177_COPY_1_24 x AI739690   ..

 Align seg 1/1  to: AI739690  from: 1  to: 277
```

```
      9 LeuLeuArgPhePheIlePheCysSerLeuAsnThrLeu 21
        |||||||||||||||:::||||||   :::|||||||:::
     19 TTGTTAAGGTTTTTCGTTTTTTGTCCCGTTAACACAATC 57

seq_name: gb_gss13:AQ562045

seq_documentation_block:
LOCUS       AQ562045      377 bp    DNA             GSS       29-MAY-1999
DEFINITION  HS_5237_A2_D10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=813 Col=20 Row=G, DNA sequence.
ACCESSION   AQ562045
VERSION     AQ562045.1  GI:4921516
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 813  row: G  column: 20
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 377.
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=813 Col=20 Row=G"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       65 a     79 c     78 g    152 t      3 others
ORIGIN


alignment_scores:
            Quality:   56.00              Length:     25
              Ratio:  3.294                Gaps:      1
 Percent Similarity:  68.000   Percent Identity: 48.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x AQ562045   ..

 Align seg 1/1  to: AQ562045  from: 1  to: 377

      6 LeuGlyPheLeuLeuArgPhePheIlePhe.................Cy 16
        |||||||||||||:::   ||||||:::|||                ||
    124 CTGGGGTTCCTGTTTGCTTTTTTCGTCTTTGAAGAGACATTTAATACTTG 173

     16 sSerLeuAsnThrLeuLeuLeuGly 24
```

```
        ||||::::::   |||:::||||||
    174 CTCAATATCTTACTTGGTATTGGGC 198

seq_name: gb_est100:BG399970

seq_documentation_block:
LOCUS       BG399970      903 bp    mRNA            EST       12-MAR-2001
DEFINITION  602442023F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4557893 5',
            mRNA sequence.
ACCESSION   BG399970
VERSION     BG399970.1  GI:13293418
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1262  row: n  column: 06
            High quality sequence stop: 573.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4557893"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      333 a    170 c    160 g    240 t
ORIGIN


alignment_scores:
           Quality:   56.00            Length:     21
             Ratio:  3.733              Gaps:      1
 Percent Similarity:  71.429   Percent Identity: 57.143

alignment_block:
 US-09-215-435-177_COPY_1_24 x BG399970/rev   ..

 Align seg 1/1  to reverse of: BG399970  from: 1  to: 903

       8 PheLeuLeuArgPhePhe.....IlePheCysSerLeuAsnTh 20
         ||||||   |||||||||          :::|||||||||||||||
     179 TTTCTTTCTAGATTTTTCTCTTTGTCTTTGGTTTTCTGCAGTTTGAATAT 130

      20 rLeuLeuLeuGly 24
         ::::::||||||
     129 GATATTTCTAGGT 117

seq_name: gb_gss37:PT010M15R

seq_documentation_block:
LOCUS       PT010M15R     983 bp    DNA             GSS       02-NOV-2000
DEFINITION  Paramecium tetraurelia sequence M10G08r of the end of plasmid
            PT010M15; contig/duplicate sequence reads: M10G08r M04D10u, genomic
            survey sequence.
ACCESSION   AL448605
VERSION     AL448605.1  GI:11122619
KEYWORDS    GSS.
SOURCE      Paramecium tetraurelia.
  ORGANISM  Paramecium tetraurelia
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
            Paramecium.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Keller,A.M. and Cohen,J.
  TITLE     An indexed genomic library for Paramecium complementation cloning
  JOURNAL   J. Eukaryot. Microbiol. 47 (1), 1-6 (2000)
  MEDLINE   20114709
REFERENCE   2  (bases 1 to 983)
  AUTHORS   Dessen,P., Zagulski,M., Gromadka,R., Plattner,H., Kissmehl,R.,
            Meyer,E., Betermier,M., Schultz,J.E., Linder,J., Pearlman,R.E.,
            Kung,C., Forney,J., Satir,B., Van Houten,J.L., Keller,A.M.,
            Froissard,M., Sperling,L. and Cohen,J.
  TITLE     Paramecium genome survey: a pilot project
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 983)
  AUTHORS   Gromadka,R. and Zagulski,M.
  TITLE     Random sequencing of the Paramecium macronuclear genome
  JOURNAL   Unpublished
  REMARK    Institute of Biochemistry and Biophysics, Polish Academy of
            Sciences, Warsaw, Poland
REFERENCE   4  (bases 1 to 983)
  AUTHORS   Cohen,J. and Sperling,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Paramecium Genome Survey Project, Centre de
            Genetique Moleculaire, Centre National de la Recherche
            Scientifique, 91198 Gif-sur-Yvette, France. E-mail:
            sperling@cgm.cnrs-gif.fr
COMMENT     The present survey of the Paramecium tetraurelia macronuclear
            genome consists of end sequences of a library of random 4-12 kb
            fragments obtained by Sau3A partial digestion of macronuclear DNA
            cloned in the BamHI site of pBSIIKS-. See [4].
            Genes are predicted from matches to other sequences. For more
            information about this sequence or the Paramecium Project, see
            http://caroll.vjf.inserm.fr/pt.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Paramecium tetraurelia"
                     /macronuclear
                     /strain="stock d4-2"
                     /db_xref="taxon:5888"
BASE COUNT      339 a    128 c     99 g    417 t
ORIGIN


alignment_scores:
           Quality:   56.00            Length:     12
             Ratio:  4.667              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x PT010M15R   ..

 Align seg 1/1  to: PT010M15R  from: 1  to: 983

       8 PheLeuLeuArgPhePheIlePheCysSerLeuAsn 19
         |||:::|||:::|||:::||||||||||||||||||
     619 TTTATTTTAAAATTTTATATCTTTTGTAGTCTAAAT 654

seq_name: gb_gss18:AQ879006

seq_documentation_block:
LOCUS       AQ879006     1524 bp    DNA             GSS       09-NOV-1999
DEFINITION  HS_3154_A2_E10_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3154 Col=20 Row=I, DNA sequence.
```

```
ACCESSION   AQ879006
VERSION     AQ879006.1  GI:6310473
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3154  row: I  column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 1524.
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3154 Col=20 Row=I"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      327 a    327 c    324 g    458 t     88 others
ORIGIN


alignment_scores:
           Quality:   56.00             Length:     18
             Ratio:  4.000               Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 61.111

alignment_block:
 US-09-215-435-177_COPY_1_24 x AQ879006   ..

 Align seg 1/1  to: AQ879006  from: 1  to: 1524

       2 LysSerAlaLysLeuGlyPheLeuLeuArgPhePheIlePheCysSerLe 18
         ||||||      ||||||||||||:::   :::|||:::|||||||||
     404 AAATCATTGTTATTGGGTTTTTTGGTTTGTTTGTTTGTTTTTTGCTCACG 453

      18 uAsn 19
         |||
     454 CAAC 457

seq_name: gb_est98:BG287285

seq_documentation_block:
LOCUS       BG287285     1047 bp    mRNA            EST       21-FEB-2001
DEFINITION  602382223F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4499909 5',
            mRNA sequence.
ACCESSION   BG287285
VERSION     BG287285.1  GI:13040973
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10364  row: n  column: 06
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4499909"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      429 a    183 c    196 g    239 t
ORIGIN


alignment_scores:
           Quality:   55.50             Length:     16
             Ratio:  3.964               Gaps:      1
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-215-435-177_COPY_1_24 x BG287285/rev   ..

 Align seg 1/1  to reverse of: BG287285  from: 1  to: 1047

       4 AlaLysLeuGlyPheLeuLeu...ArgPhePheIlePheCysSerLeu 18
         |||   ||||||||||||||   |||||||||:::|||||||||:::
     873 GCGCCCCTTGGTTTTTTGTTGCCTCGCTTTTTTTTGTTTTGTTCTGTT 826
```

OM of: US-09-215-435-177_COPY_25_99 to: GenEmbl:* out_format : pfs

Date: Nov 15, 2001 12:41 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_pr9:HSDJ601O1	+	341.00	675.46	2.2e-29	33458	! AL109656 Human DNA sequence f
gb_htg9:AC023971	-	341.00	662.63	1.1e-28	153788	! AC023971 Homo sapiens chromc
gb_om:BTU35642	+	134.00	268.46	1.0e-06	1198	! U35642 Bos taurus alpha1-micrc
gb_in3:CELC37C3	+	133.00	236.34	6.3e-05	42521	! U64857 Caenorhabditis elegans
gb_htg1:AC006915	+	133.00	226.68	0.0002	134095	! AC006915 Caenorhabditis eleg
gb_htg6:AC019993	-	131.00	231.32	0.0001	46847	! AC019993 Drosophila melanogas
gb_in1:AC003053	+	131.00	226.75	0.0002	80689	! AC003053 Drosophila melanogas
gb_in1:AC005149	-	131.00	225.04	0.0003	98942	! AC005149 Drosophila melanogas
gb_in1:AE003579	-	131.00	215.59	0.0009	304383	! AE003579 Drosophila melanoga
gb_om:MEELACP	+	130.00	268.88	9.7e-07	419	! AJ000490 Macropus eugenii mRNA
gb_pat2:I63559	+	130.00	258.42	3.7e-06	1455	! I63559 Sequence 6 from patent
gb_pr9:HSAIMBG3	+	129.50	247.64	1.5e-05	4627	! X54818 Human gene for alpha-1-
gb_pr10:HUMITILC08	+	129.50	246.30	1.8e-05	5423	! M88249 Human inter-alpha-tryps
gb_htg19:AL137850	-	129.50	216.22	0.0008	193939	! AL137850 Homo sapiens chromc
gb_htg3:AC011161	-	129.50	215.94	0.0009	200532	! AC011161 Homo sapiens clone
gb_sy:SYNBOVTRYI	+	129.00	269.56	8.9e-07	301	! M31335 Bovine synthetic colostr
gb_in2:AF078161	+	129.00	239.30	4.3e-05	11008	! AF078161 Manduca sexta lacuni
gb_in2:CBRG45N02	+	129.00	227.89	0.0002	42726	! AC084649 Caenorhabditis brigg
gb_pat2:E12656	+	128.50	263.37	2.0e-06	555	! E12656 Synthetic DNA encoding u
gb_pr8:HSA1MICR	+	128.50	256.74	4.6e-06	1221	! X04494 Human mRNA for alpha-1-
gb_pat2:I05217	+	128.50	256.66	4.6e-06	1232	! I05217 Sequence 7 from Patent
gb_pr9:HSHCR	+	128.50	256.62	4.7e-06	1239	! X04225 Human mRNA for protein
gb_pat1:A31024	+	127.00	269.19	9.3e-07	191	! A31024 DNA fragment for bikunin
gb_pat1:A31025	-	127.00	269.19	9.3e-07	191	! A31025 DNA fragment for bikunin
gb_pat1:A19212	+	127.00	262.60	2.2e-06	418	! A19212 Human kunitz type protea
gb_pat2:I29002	+	127.00	262.60	2.2e-06	418	! I29002 Sequence 17 from patent
gb_om:TVU34208	+	126.00	261.35	2.5e-06	378	! U34208 Trichosurus vulpecula ea
gb_pat2:E13092	+	124.00	255.96	5.1e-06	435	! E13092 DNA encoding a protease
gb_ro1:MMAMBP	+	124.00	247.57	1.5e-05	1180	! X68680 M.musculus AMBP mRNA fc
gb_ro1:MUSAMIATI	+	124.00	247.19	1.6e-05	1234	! D28812 Mouse mRNA for alpha-1
gb_pr9:HSDJ447F3	+	124.00	210.46	0.0017	97385	! AL050348 Human DNA sequence f
gb_om:OAU00163	+	123.50	262.62	2.2e-06	174	! U00163 Ovis aries secretory prc
gb_om:AF241828	+	123.50	247.12	1.6e-05	1098	! AF241828 Ovis aries kunitz dom
gb_pat1:A19210	+	123.00	254.19	6.4e-06	418	! A19210 Human kunitz type protea
gb_pat2:I29001	+	123.00	254.19	6.4e-06	418	! I29001 Sequence 15 from patent
gb_in3:CEW01F3	+	123.00	217.51	0.0007	32797	! Z92815 Caenorhabditis elegans
gb_htg24:CEY44A6	+	123.00	198.19	0.0084	326074	! Z98863 Caenorhabditis elegan
gb_ro2:S87544	+	122.00	243.50	2.5e-05	1162	! S87544 polyprotein 1-microglob
gb_ov:AF008648	+	122.00	240.51	3.7e-05	1657	! AF008648 Cyprinus carpio unknc
gb_pr5:AK024817	+	121.00	245.33	2.0e-05	728	! AK024817 Homo sapiens cDNA: FLJ
gb_om:S61902	+	121.00	243.20	2.6e-05	938	! S61902 tissue factor pathway in
gb_pat2:I14875	+	121.00	242.84	2.7e-05	979	! I14875 Sequence 1 from paten
gb_pat2:I92685	+	121.00	242.84	2.7e-05	979	! I92685 Sequence 1 from paten
gb_pr10:HUMHTFP	+	121.00	242.84	2.7e-05	979	! L27624 Homo sapiens tissue f
gb_om:OCLIPOPR	+	121.00	242.17	3.0e-05	1060	! X54708 Rabbit mRNA for lipo

seq_name: gb_pr9:HSDJ601O1

seq_documentation_block:

LOCUS HSDJ601O1 33458 bp DNA PRI 04-APR-2001

DEFINITION Human DNA sequence from clone RP4-601O1 on chromosome 20. Contains the gene for a novel protein with Kunitz/Bovine pancreatic trypsin inhibitor domain, the 5' end of a novel gene, part of the gene for a novel protein with WAP-type (Whey Acidic Protein) 'four-disulfide core' domain, ESTs and GSSs, complete sequence.

ACCESSION AL109656

VERSION AL109656.10 GI:10944958

KEYWORDS HTG; Kunitz; WAP.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 33458)

AUTHORS Laird,G.

TITLE Direct Submission

JOURNAL Submitted (03-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Oct 21, 2000 this sequence version replaced gi:9795206.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone RP4-601O1 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP3-447F3 is at 33359 in this sequence.
The true right end of clone RP4-688G8 is at 100 in this sequence.
RP4-601O1 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES Location/Qualifiers

source 1. .33458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP4-601O1"
/clone_lib="RPCI-4"

repeat_region 3. .225
/note="L1M4 repeat: matches 3977. .4203 of consensus"

repeat_region 225. .428
/note="L1M4 repeat: matches 4321. .4525 of consensus"

repeat_region 429. .696
/note="AluJb repeat: matches 2. .302 of consensus"

repeat_region 697. .778

```
                /note="L1M4 repeat: matches 4525. .4606 of consensus"
repeat_region   779. .1089
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   1090. .1146
                /note="L1M4 repeat: matches 4606. .4662 of consensus"
repeat_region   1165. .1293
                /note="LTR42 repeat: matches 1. .134 of consensus"
repeat_region   1351. .1730
                /note="L1MB5 repeat: matches 5005. .5402 of consensus"
repeat_region   1731. .2032
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   2056. .2547
                /note="L1MB5 repeat: matches 5374. .5865 of consensus"
repeat_region   2543. .2729
                /note="L1MC/D repeat: matches 5428. .5609 of consensus"
repeat_region   2730. .3022
                /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region   3023. .3181
                /note="L1MC/D repeat: matches 5609. .5769 of consensus"
repeat_region   3361. .3895
                /note="L1 repeat: matches 3599. .4167 of consensus"
repeat_region   3986. .4281
                /note="L2 repeat: matches 534. .834 of consensus"
repeat_region   4526. .4828
                /note="AluSx repeat: matches 1. .300 of consensus"
misc_feature    5372. .5726
                /note="match: GSS: Em:AQ076126"
repeat_region   5394. .5470
                /note="MIR repeat: matches 48. .142 of consensus"
gene            7401. .7609
                /gene="dJ601O1.3"
CDS             <7401. .>7609
                /gene="dJ601O1.3"
                /note="possibly a pseudogene
                match: proteins: Sw:Q14508 Sw:P00993 Sw:Q28631 Sw:P14730
                Tr:O44131 Wp:C08G9"
                /codon_start=2
                /evidence=not_experimental
                /product="dJ601O1.3 (novel protein with WAP-type (Whey
                Acidic Protein) 'four-disulfide core' domain)"
                /protein_id="CAC36024.1"
                /db_xref="GI:13559237"
                /translation="FPGSSLGFLSSEHYSSLLVRPFSYLFCVPEYILEPPPCISAPEN
                CTHLCTMQEDCEKGFQCCSSFCGIV"
mRNA            complement(7623. .7780)
                /gene="dJ688G8.4"
                /note="Continues in Em:AL031671 as dJ688G8.4
                match: ESTs: Em:AI808149 Em:AI808159"
                /evidence=not_experimental
                /product="dJ601O1.2 (a putative novel protein)"
gene            complement(7623. .7780)
                /gene="dJ688G8.4"
repeat_region   8352. .8422
                /note="L2 repeat: matches 2596. .2670 of consensus"
repeat_region   8434. .8656
                /note="MIR repeat: matches 9. .260 of consensus"
repeat_region   9290. .9419
                /note="MIR repeat: matches 47. .183 of consensus"
misc_feature    complement(9433. .9899)
                /note="match: GSS: Em:AQ476416"
repeat_region   9538. .9830
                /note="AluSg repeat: matches 1. .293 of consensus"
misc_feature    9923. .10651
                /note="match: GSS: Em:AQ478366"
repeat_region   10453. .10647
                /note="MER3 repeat: matches 1. .204 of consensus"
misc_feature    complement(10590. .10982)
                /note="match: GSS: Em:AQ223524"
repeat_region   11007. .11080
                /note="MER69 repeat: matches 2. .77 of consensus"
repeat_region   11529. .11954
                /note="MER69 repeat: matches 2070. .2511 of consensus"
repeat_region   13092. .13158
                /note="L2 repeat: matches 2666. .2732 of consensus"
repeat_region   13202. .13331
                /note="AluSg/x repeat: matches 183. .312 of consensus"
repeat_region   13332. .15510
                /note="L1MB1 repeat: matches 3916. .6171 of consensus"
repeat_region   15534. .15832
                /note="AluSg repeat: matches 1. .299 of consensus"
repeat_region   15846. .16746
                /note="L1M4 repeat: matches 3472. .4394 of consensus"
repeat_region   16739. .18594
                /note="L1PA11 repeat: matches 11. .1921 of consensus"
repeat_region   18595. .18888
                /note="AluSq repeat: matches 1. .295 of consensus"
repeat_region   18889. .20972
                /note="L1PA11 repeat: matches 1921. .3456 of consensus"
repeat_region   20973. .21280
                /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region   21281. .23911
                /note="L1PA11 repeat: matches 3456. .6164 of consensus"
repeat_region   23926. .24129
                /note="L1M4 repeat: matches 4390. .4613 of consensus"
repeat_region   24130. .24438
                /note="AluSx repeat: matches 1. .310 of consensus"
repeat_region   24439. .24513
                /note="L1M4 repeat: matches 4613. .4683 of consensus"
repeat_region   24790. .25008
                /note="MIR repeat: matches 1. .242 of consensus"
mRNA            join(25397. .25528,26926. .27103,28675. .28876)
                /gene="dJ601O1.1"
                /note="match: cDNAs: Em:AK020315
                match: ESTs: Em:AW303354"
                /evidence=not_experimental
                /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                pancreatic trypsin inhibitor domain)"
gene            25397. .28876
                /gene="dJ601O1.1"
CDS             join(<25397. .25528,26926. .27103,28675. .28681)
                /gene="dJ601O1.1"
                /note="match: proteins: Sw:P08592 Tr:Q13793 Sw:P05067
                Sw:P12023 Sw:Q95241 Tr:Q9TUI0 Tr:Q91963 Sw:Q02445
                Tr:O54819 Tr:Q9Z2U8 Sw:P12111 Tr:Q9XWX5 Tr:O61893
                Sw:P00974 Sw:P04815 Tr:O16701 Sw:P00975 Tr:O70160
                Sw:P02760 Sw:P48307 Wp:C37C3A Sw:P00991 Sw:P00978
                Sw:P00976 Sw:P13371 Sw:P04365 Tr:Q9VQT9 Tr:O95925
                Sw:P00993 Sw:P10280 Tr:Q9TWF8 Tr:Q9TWG0 Tr:Q9TWF9
                Sw:P26228 Sw:P00980 Sw:P00992 Sw:P00979 Sw:P81658
                Sw:P25660 Sw:P19761 Sw:P24541 Tr:Q9U8G8 Tr:Q14594
                Tr:O95428 Tr:O35536"
                /codon_start=3
                /evidence=not_experimental
                /product="dJ601O1.1 (novel protein with Kunitz/Bovine
                pancreatic trypsin inhibitor domain)"
                /protein_id="CAC36025.1"
                /db_xref="GI:13559238"
                /translation="AACCTMKSAKLGFLLRFFIFCSLNTLLLGGVNKIAEKICGDLKD
                PCKLDMNFGSCYEVHFRYFYNRTSKRCETFVFSGCNGNLNNFKLKIEREVACVAKYKP
                PR"
repeat_region   26066. .26228
                /note="MIR repeat: matches 53. .205 of consensus"
repeat_region   27253. .27579
                /note="AluSp repeat: matches 1. .303 of consensus"
repeat_region   27649. .27728

alignment_scores:
            Quality:  341.00              Length:     60
              Ratio:  5.683                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_25_99 x HSDJ601O1   ..

 Align seg 1/1  to: HSDJ601O1  from: 1  to: 33458
```

```
    15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
 26925 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 26974

    31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
 26975 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 27024

    48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
 27025 CCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 27074

    65 ValAlaCysValAlaLysTyrLysProPro 74
       ||||||||||||||||||||||||||||||
 27075 GTAGCCTGTGTTGCAAAATACAAACCACCG 27104

seq_name: gb_htg9:AC023971

seq_documentation_block:
LOCUS       AC023971   153788 bp    DNA             HTG       07-APR-2000
DEFINITION  Homo sapiens chromosome 20 clone RP11-770N8 map 20, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC023971
VERSION     AC023971.2  GI:7523758
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 20, clone RP11-770N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 7, 2000 this sequence version replaced gi:7008875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6573
                Center clone name: 770_N_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145038 bases at least Q40
                Consensus quality: 148490 bases at least Q30
                Consensus quality: 150621 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 152588; sum-of-contigs
                Quality coverage: 5.2 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1333: contig of 1333 bp in length
            *     1334 1433: gap of      100 bp
            *     1434     3969: contig of 2536 bp in length
            *     3970 4069: gap of      100 bp
            *     4070     7211: contig of 3142 bp in length
            *     7212 7311: gap of      100 bp
            *     7312    13518: contig of 6207 bp in length
            *    13519 13618: gap of      100 bp
            *    13619    20876: contig of 7258 bp in length
            *    20877 20976: gap of      100 bp
            *    20977    28838: contig of 7862 bp in length
            *    28839 28938: gap of      100 bp
            *    28939    37590: contig of 8652 bp in length
            *    37591 37690: gap of      100 bp
            *    37691    45814: contig of 8124 bp in length
            *    45815 45914: gap of      100 bp
            *    45915    62601: contig of 16687 bp in length
            *    62602 62701: gap of      100 bp
            *    62702    81817: contig of 19116 bp in length
            *    81818 81917: gap of      100 bp
            *    81918   100273: contig of 18356 bp in length
            *   100274 100373: gap of      100 bp
            *   100374   125266: contig of 24893 bp in length
            *   125267 125366: gap of      100 bp
            *   125367   153788: contig of 28422 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20"
                     /clone="RP11-770N8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .3969
                     /note="assembly_fragment"
     misc_feature    4070. .7211
                     /note="assembly_fragment"
     misc_feature    7312. .13518
                     /note="assembly_fragment"
     misc_feature    13619. .20876
                     /note="assembly_fragment"
     misc_feature    20977. .28838
                     /note="assembly_fragment"
     misc_feature    28939. .37590
                     /note="assembly_fragment"
     misc_feature    37691. .45814
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    45915. .62601
                     /note="assembly_fragment"
     misc_feature    62702. .81817
                     /note="assembly_fragment
                     clone_end:T7
```

```
                     vector_side:right"
     misc_feature    81918. .100273
                     /note="assembly_fragment"
     misc_feature    100374. .125266
                     /note="assembly_fragment"
     misc_feature    125367. .153788
                     /note="assembly_fragment"
BASE COUNT    44583 a   30904 c   30568 g   46532 t    1201 others
ORIGIN


alignment_scores:
            Quality:  341.00              Length:      60
              Ratio:  5.683                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_25_99 x AC023971/rev   ..

 Align seg 1/1  to reverse of: AC023971  from: 1  to: 153788

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  134123 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 134074

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  134073 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 134024

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  134023 CCGGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 133974

      65 ValAlaCysValAlaLysTyrLysProPro 74
         |||||||||||||||||||||||||||||||
  133973 GTAGCCTGTGTTGCAAAATACAAACCACCG 133944

seq_name: gb_om:BTU35642

seq_documentation_block:
LOCUS       BTU35642     1198 bp    mRNA            MAM       05-SEP-1996
DEFINITION  Bos taurus alpha1-microglobulin/bikunin mRNA, complete cds.
ACCESSION   U35642
VERSION     U35642.1  GI:1016297
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1198)
  AUTHORS   Lindqvist,A. and Akerstrom,B.
  TITLE     Bovine alpha 1-microglobulin/bikunin. Isolation and
            characterization of liver cDNA and urinary alpha 1-microglobulin
  JOURNAL   Biochim. Biophys. Acta, Gene Struct. Expr. 1306 (1), 98-106 (1996)
  MEDLINE   96201710
REFERENCE   2  (bases 1 to 1198)
  AUTHORS   Lindqvist,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1995) Annika Lindqvist, Molecular Signalling,
            Lund University, P.O. Box 94, Lund, 221 00, Sweden
FEATURES             Location/Qualifiers
     source          1. .1198
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="liver"
     CDS             104. .1162
                     /note="precursor protein for alpha1-microglobulin and
                     bikunin"
                     /codon_start=1
                     /product="alpha1-microglobulin/bikunin"
                     /protein_id="AAB07599.1"
                     /db_xref="GI:1016298"
                     /translation="MRSLSGLLLLLTACLAVNASSVPTLPDDIQVQENFDLSRIYGKW
                     FNVAVGSTCPWLKRFKEKMTMSTVVLIAGPTSKEISVTNTHRRKGVCESISGTYEKTS
                     ADGKFLYHKAKWNITMESYVVHTNYDEYAIFLTKKLSRRHGPTITVKLYGREPQLRES
                     LLEEFREVALGVGIPEDAIFTMPDRGECVPGEQDPVPTPLSRARRAVLTQEEEGSGAG
                     QPVTNFSKKADSCQLDYSQGPCLGLFKRYFYNGTSMACETFLYGGCMGNGNNFLSEKE
                     CLQTCRTVEACNLPIVQGPCRSYIQLWAFDAVKGKCVRFSYGGCKGNGNKFYSEKECK
                     EYCGIPGEADEELLRFSN"
BASE COUNT      284 a     308 c     354 g     252 t
ORIGIN


alignment_scores:
            Quality:  134.00              Length:      43
              Ratio:  4.467                Gaps:       0
 Percent Similarity:  69.767   Percent Identity: 55.814

alignment_block:
 US-09-215-435-177_COPY_25_99 x BTU35642   ..

 Align seg 1/1  to: BTU35642  from: 1  to: 1198

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||:::||||||   :::   |||   |||      :::
     788 GATTCCTGCCAGCTGGACTACTCACAAGGGCCTTGCCTGGGGCTGTTCAA 837

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||   ||||||      ||||||||||||:::::::
     838 GAGGTATTTCTATAACGGTACATCCATGGCCTGTGAGACCTTCTTATATG 887

      48 erGlyCysAsnGlyAsnLeuAsnAsnPhe 57
         ::|||||   ||||||   |||||||||
     888 GCGGCTGCATGGGGAATGGCAACAACTTC 916

seq_name: gb_in3:CELC37C3

seq_documentation_block:
LOCUS       CELC37C3    42521 bp    DNA             INV       08-JUL-1998
DEFINITION  Caenorhabditis elegans cosmid C37C3.
ACCESSION   U64857
VERSION     U64857.1  GI:1465829
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 42521)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 42521)
  AUTHORS   Geisel,C. and Bradshaw,H.
  TITLE     The sequence of C. elegans cosmid C37C3
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 42521)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-1996)
REFERENCE   4  (bases 1 to 42521)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1998) Department of Genetics, Washington
```

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is ZK742, 200 bp overlap;3' cosmid is F40A3, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELC37C3; actual end is at 5024 of CELF40A3

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers
source 1. .42521
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="C37C3"
/chromosome="V"
gene 513. .15440
/gene="C37C3.6b"
gene 513. .6558
/gene="C37C3.6a"
CDS join(513. .608,664. .878,1380. .1535,1583. .1734,1785. .2541,
2666. .2897,2976. .3097,3149. .4733,4801. .5584,5659. .5871,
6194. .6558)
/gene="C37C3.6a"
/note="similar to the DPTI/Kunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor; coded for by C. elegans cDNA yk84d8.5; coded for by C. elegans cDNA yk117d1.5; coded for by C. elegans cDNA yk88e5.3; coded for by C. elegans cDNA yk132d4.3; coded for by C. elegans cDNA yk123e8.3; coded for by C. elegans cDNA yk88e5.5; coded for by C. elegans cDNA yk22b2.5; coded for by C. elegans cDNA yk132d4.5; coded for by C. elegans cDNA yk123e8.5; coded for by C. elegans cDNA CEESD32F; coded for by C. elegans cDNA CEESH31FC; coded for by C. elegans cDNA CEESL67F; coded for by C. elegans cDNA CEESL73F; coded for by C. elegans cDNA CEESP33F; coded for by C. elegans cDNA CEESP94F; coded for by C. elegans cDNA CESAC89F; coded for by C. elegans cDNA yk10f9.3; coded for by C. elegans cDNA yk10f9.5; coded for by C. elegans cDNA yk22b2.3; coded for by C. elegans cDNA yk3e8.3; coded for by C. elegans cDNA yk3e8.5; coded for by C. elegans cDNA yk41h8.5; coded for by C. elegans cDNA yk5c8.5; coded for by C. elegans cDNA yk148d2.5; coded for by C. elegans cDNA yk152f10.5"
/codon_start=1
/protein_id="AAC25867.1"
/db_xref="GI:3294501"
/translation="MRLLLFSAALLLCSVPTWAFSLSSFFGSDVAQKPYLHPNSPPER
DPASSRMKRQAYQVYVDGDVSVTVDKSGQKETGNWGPWVPENECSRSCGGGVQLEKRQ
CSGDCTGASVRYISCNLNACESGTDFRAEQCSKFNDEALDGNYHKWTPYKGKNKCELV
CKPESGNFYYKWADKVVDGTKCDSKSNDICVDGECLPVGCDGKLGSSLKFDKCGKCDG
DGSTCKTIEGRFDERNLSPGYHDIIKLPEGATNIKIQEARKSTNNLALKNGSDHFYLN
GNGLIQVEKEVEVGGTIFVYDDAEPETLSAQGPLSEELTVALLFRKGSRDTAIKYEFS
IPLEEEVDYMYKFDNWTPCSVSCGKGVQTRNLYCIDGKNKGRVEDDLCEENNATKPEF
EKSCETVDCEAEWFTGDWESCSSTCGDQGQQYRVVYCHQVFANGRRVTVEDGNCTVER
PPVKQTCNRFACPEWQAGPWSACSEKCGDAFQYRSVTCRSEKEGEEGKLLAADACPAD
EQEKFDTERTCNLGPCEGLTFVTGEWNLCTRCNDTEETREVTCKDSQGRAYPLEKCLV
DNSTEIPTDTRSCATQPPCEYEWTVSEWSKCTTECGHGHKTRRVICAIHQNGGLEVVD
EGHCQAEKPEGKTNCTNEEKCTGTWYTSSWSECTAECGGGSQDRVAVCLNYDKKPVPE
WCDEAVKPSEKQDCNVDDCPTCVDSEFGCCPDNSTFATGEFNFGCSNCSETEFGCCAD
NVTVATGPNSKGCEEFVESPLNLEADVANADAEASGDAPELCSVTNENGEAVDVECAT
IAPITALLGDGELIGNDTDASNETIHCSKTEFGCCPDWYTAASGKGNEGCPSFTLGGC
NETQFGCCHDDVTLARGANLEGCGEPSCAASLYGCCKDRKTIAFGPHYSGCERSSFPC
ELSDFGCCPDGETAALGKNGTGCGENCLTTKFGCCPDGKTTAKGSHNEGCGCEFAQYG
CCPDGKSVAKGAGFYGCPESCAQSQFGCCPDGKTRARGENKEGCPCQYTRYGCCPDGE
TTALGPRNDGCDNCRYAKHGCCPDGETKALGPDGAGCPPTTTPPFLMGGTVAPHKIAA
CNQTQESGTVCGAGYKLAWHYDTTEGRCNQFWYGGCGGNDNNFASQDMCETICVEPPG
KGRCYLPRVDGPLRCDQLQPRYYYDHSKKHCVAFWWRGCLGNANNFNSFEECSMFCKD
VGPYDAPTTAAPPPPPQQNAQQYLPTPEVQQIEIQSAEQPQPQQPQQQQQQQQQQPQQ
PRQSMEDICRSRQDAGPCETYSDQWFYNAFSQECETFTYGGCGGNLNRFRSKDECEQR
CFFVHGAQPSAARQEQAQPAAQPAQPAQPSNIVSPPQQSASPVVVPSNSKQRDACHLN
VDQGRCKGAFDSWYYEVATGSCVTFKYTGCGGNANRFASKDQCESLCVKPASEAASAG
IDGAAGINSVCDEAKDTGPCTNFVTKWYYNKADGTCNRFHYGGCQGTNNRFDNEQQCK
AACQNHKDACQLPKVQGPCSGKHSYYYYNTASHQCETFTYGGCLGNTNRFATIEECQA
RCPSKF"
CDS join(513. .608,664. .878,1380. .1535,1583. .1734,1785. .2541,
2666. .2897,2976. .3097,3149. .4733,4801. .5584,5659. .5871,
6194. .6547,7038. .7134,7881. .8009,8060. .8410,8493. .8571,
13650. .13741,14139. .14727,14940. .15440)
/gene="C37C3.6b"
/note="similar to the BPTI/Kunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor (TFPI); coded for by C. elegans cDNA yk257g8.5; coded for by C. elegans cDNA yk257g8.3; coded for by C. elegans cDNA yk185b12.3"
/codon_start=1
/protein_id="AAC25868.1"
/db_xref="GI:3294502"
/translation="MRLLLFSAALLLCSVPTWAFSLSSFFGSDVAQKPYLHPNSPPER
DPASSRMKRQAYQVYVDGDVSVTVDKSGQKETGNWGPWVPENECSRSCGGGVQLEKRQ
CSGDCTGASVRYISCNLNACESGTDFRAEQCSKFNDEALDGNYHKWTPYKGKNKCELV
CKPESGNFYYKWADKVVDGTKCDSKSNDICVDGECLPVGCDGKLGSSLKFDKCGKCDG
DGSTCKTIEGRFDERNLSPGYHDIIKLPEGATNIKIQEARKSTNNLALKNGSDHFYLN
GNGLIQVEKEVEVGGTIFVYDDAEPETLSAQGPLSEELTVALLFRKGSRDTAIKYEFS
IPLEEEVDYMYKFDNWTPCSVSCGKGVQTRNLYCIDGKNKGRVEDDLCEENNATKPEF
EKSCETVDCEAEWFTGDWESCSSTCGDQGQQYRVVYCHQVFANGRRVTVEDGNCTVER
PPVKQTCNRFACPEWQAGPWSACSEKCGDAFQYRSVTCRSEKEGEEGKLLAADACPAD
EQEKFDTERTCNLGPCEGLTFVTGEWNLCTRCNDTEETREVTCKDSQGRAYPLEKCLV
DNSTEIPTDTRSCATQPPCEYEWTVSEWSKCTTECGHGHKTRRVICAIHQNGGLEVVD
EGHCQAEKPEGKTNCTNEEKCTGTWYTSSWSECTAECGGGSQDRVAVCLNYDKKPVPE
WCDEAVKPSEKQDCNVDDCPTCVDSEFGCCPDNSTFATGEFNFGCSNCSETEFGCCAD
NVTVATGPNSKGCEEFVESPLNLEADVANADAEASGDAPELCSVTNENGEAVDVECAT
IAPITALLGDGELIGNDTDASNETIHCSKTEFGCCPDWYTAASGKGNEGCPSFTLGGC
NETQFGCCHDDVTLARGANLEGCGEPSCAASLYGCCKDRKTIAFGPHYSGCERSSFPC
ELSDFGCCPDGETAALGKNGTGCGENCLTTKFGCCPDGKTTAKGSHNEGCGCEFAQYG
CCPDGKSVAKGAGFYGCPESCAQSQFGCCPDGKTRARGENKEGCPCQYTRYGCCPDGE
TTALGPRNDGCDNCRYAKHGCCPDGETKALGPDGAGCPPTTTPPFLMGGTVAPHKIAA
CNQTQESGTVCGAGYKLAWHYDTTEGRCNQFWYGGCGGNDNNFASQDMCETICVEPPG
KGRCYLPRVDGPLRCDQLQPRYYYDHSKKHCVAFWWRGCLGNANNFNSFEECSMFCKD
VGPYDAPTTAAPPPPQQNAQQYLPTPEVQQIEIQSAEQPQPQQPQQQQQQQQQQPQQ
PRQSMEDICRSRQDAGPCETYSDQWFYNAFSQECETFTYGGCGGNLNRFRSKDECEQR
CFFVHGAQPSAARQEQAQPAAQPAQPAQPSNIVSPPQQSASPVVVPSNSKQRDACHLN
VDQGRCKGAFDSWYYEVATGSCVTFKYTGCGGNANRFASKDQCESLCVKPASEAASAG
IDGAAGINSVCDEAKDTGPCTNFVTKWYYNKADGTCNRFHYGGCQGTNNRFDNEQQCK
AACQNHKDACQLPKVQGPCSGKHSYYYYNTASHQCETFTYGGCLGNTNRFATIEECQA
RCPKDDQTTTTSQPEELPSLPLVQEDPQPRPAFSLKQSFAHSRRRDAPFARSVSARHH
TPDSEEERVDCYAVPDPGSCGDYRLVWHYSATSNSCRQFYYGGCAGNTNRFETRDKCE
TSCVAKIEERVESVSEASKSLEEVRLTDPRMDSHFGYHDPEVDQIEEEAEYVIVDTGA
LPELCMLPEQRGSCYDNILRWRFDSEKSQCVTFMYSGCNPNANHFTSQETCERACGKW
RNVAVCELPAEHGDCQLAIPRWYHDPKTSQCQMMMWTGCGGNGNAFSSKADCESLCRV
ETLWSNNTDFCTLERSAGPCTDSISMWYFDSTHLDCKPFTYGGCRGNQNRFVSKEQCQ
QSCRPGDTKSEDICTLRPEPGPCRLGLEKYFYDPVIQSCHMFHYGGCEGNANRFDSEL

```
                DCFRRCSSVKVEASESERVGQLTSASTPVIYIVNKTAIFVGNTFRIRCNSYGVLPITW
                YKNGGLLQFGSRITEENDDTLEIVDALTADAGVYTCIAGQDSTMSEGVEVVIKRLPGH
                RTTSRPMLTPSKNFSLGTPPTPSPSTVSTTPFRIYTPGSAPSDARVSRPTSNSCMDVG
                NASTCDLIVKNGLCGKKRYGTFCCHTCTRVHNFKF"
     gene       complement(8582. .12734)
                /gene="C37C3.7"
     CDS        complement(join(8582. .8744,8795. .8868,8914. .9194,
                9378. .9558,9607. .9722,10028. .10144,10473. .10730,
                10960. .11122,11912. .12280,12651. .12734))
                /gene="C37C3.7"
                /note="contains similarity to EGF-like domains
                (PS:PS00022)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC25862.1"
                /db_xref="GI:1465832"
                /translation="MTWSEKQEYCRIGSNDLNTCETCVGKGSNCFWCGGKTKRCMPFD

alignment_scores:
            Quality:  133.00              Length:     62
              Ratio:  3.500                 Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177_COPY_25_99 x CELC37C3    ..

 Align seg 1/1  to: CELC37C3  from: 1  to: 42521

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   :::   ||||||   |||:::|||            |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |        |||      |||:::|||||||   :::|||   :::||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      43 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 59
         ||||||||:::::::::|||||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      60 LysIleGluArgGluValAlaCysValAlaLysTyr 71
            |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555

 seq_name: gb_htg1:AC006915

 seq_documentation_block:
 LOCUS       AC006915   134095 bp    DNA              HTG       24-FEB-1999
 DEFINITION  Caenorhabditis elegans clone Y97E10y, *** SEQUENCING IN PROGRESS
             ***, 1 unordered pieces.
 ACCESSION   AC006915
 VERSION     AC006915.1  GI:4263433
 KEYWORDS    HTG; HTGS_PHASE1.
 SOURCE      Caenorhabditis elegans.
   ORGANISM  Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE   1  (bases 1 to 134095)
   AUTHORS   Waterston,R.H.
   TITLE     The sequence of Caehorhabditis elegans clone
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 134095)
   AUTHORS   Waterston,R.H.
   TITLE     Direct Submission
   JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
 COMMENT     * NOTE: This is a 'working draft' sequence. It currently
             * consists of 1 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1   134095: contig of 134095 bp in length.
 FEATURES             Location/Qualifiers
      source          1. .134095
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /clone="Y97E10y"
 BASE COUNT    43230 a   24499 c   23702 g   42664 t
 ORIGIN


 alignment_scores:
             Quality:  133.00              Length:     62
               Ratio:  3.500                 Gaps:      0
  Percent Similarity:  61.290   Percent Identity: 40.323

 alignment_block:
  US-09-215-435-177_COPY_25_99 x AC006915    ..

  Align seg 1/1  to: AC006915  from: 1  to: 134095

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   :::   ||||||   |||:::|||            |||   ||
   54670 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 54719

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |        |||      |||:::|||||||   :::|||   :::||||
   54720 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 54769

      43 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 59
         ||||||||:::::::::|||||||   ||||||   |||:::|||
   54770 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 54819

      60 LysIleGluArgGluValAlaCysValAlaLysTyr 71
            |||   ::::::   |||   :::|||:::
   54820 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 54855

 seq_name: gb_htg6:AC019993

 seq_documentation_block:
 LOCUS       AC019993    46847 bp    DNA              HTG       03-JAN-2000
 DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
             pieces.
 ACCESSION   AC019993
 VERSION     AC019993.1  GI:6664904
 KEYWORDS    HTG; HTGS_PHASE2.
 SOURCE      fruit fly.
   ORGANISM  Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE   1  (bases 1 to 46847)
   AUTHORS   Adams,M. and Venter,J.C.
   TITLE     Direct Submission
   JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
             Rockville, MD, USA
 COMMENT     This sequence was identified as CDM:10211667 by the submitter.
             For more information on this record e-mail to fly@celera.com.
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
 FEATURES             Location/Qualifiers
      source          1. .46847
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
 BASE COUNT    13884 a   9574 c   9324 g   14065 t
 ORIGIN


 alignment_scores:
             Quality:  131.00              Length:     56
```

```
            Ratio:  3.639               Gaps:      0
 Percent Similarity:  64.286   Percent Identity: 41.071

alignment_block:
 US-09-215-435-177_COPY_25_99 x AC019993/rev   ..

 Align seg 1/1  to reverse of: AC019993  from: 1  to: 46847

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         :::   |||            :::   |||   |||:::   :::   ::
    5598 GAGGACTGTCATCAGCCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTA 5549

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         :||||||   ||||||      ::::::   ||||||   ||||||::::
    5548 TCGCTACGCCTACAACGTGGATACACAATCCTGCGAGGAGTTCGTTTACG 5499

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||   |||||||||:::   |||   :::   |||
    5498 GTGGATGTGCCGGCAACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAA 5449

      65 ValAlaCysValAlaLys 70
         ||||||::::::|||
    5448 CAGGCATGTTTGGTAAAG 5431

seq_name: gb_in1:AC003053

seq_documentation_block:
LOCUS       AC003053    80689 bp    DNA             INV       17-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS00216 (D160)), complete
            sequence.
ACCESSION   AC003053 AC002863 AC003283 AC003284 AC003285 AC003286 AC003287
            AC002864 AC003288 AC003289 AC003290 AC003291 AC003292 AC003293
            AC003294 AC003295 AC003296 AC003297
VERSION     AC003053.1  GI:2584830
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clone DS00216
            (D160)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 80689)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 24A3-24C2
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 80689)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library location: 24-3.

FEATURES             Location/Qualifiers
     source          1. .80689
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="24A3-24C2"
                     /clone="P1 DS00216 (D160)"
                     /note="This sequence has not changed since its original
                     submission on 11/04/1997.  It was resubmitted in order to
                     include all secondary accession numbers for the subclones
                     belonging to this clone."
BASE COUNT    23494 a  16620 c  16241 g  24334 t
ORIGIN


alignment_scores:
            Quality:  131.00             Length:     56
              Ratio:  3.639                Gaps:      0
 Percent Similarity:  64.286   Percent Identity: 41.071

alignment_block:
 US-09-215-435-177_COPY_25_99 x AC003053   ..

 Align seg 1/1  to: AC003053  from: 1  to: 80689

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         :::   |||            :::   |||   |||:::   :::   ::
    6096 GAGGACTGTCATCAGCCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTA 6145

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         :||||||   ||||||      ::::::   ||||||   ||||||::::
    6146 TCGCTACGCCTACAACGTGGATACACAATCCTGCGAGGAGTTCGTTTACG 6195

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||   |||||||||:::   |||   :::   |||
    6196 GTGGATGTGCCGGCAACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAA 6245

      65 ValAlaCysValAlaLys 70
         ||||||::::::|||
    6246 CAGGCATGTTTGGTAAAG 6263

seq_name: gb_in1:AC005149

seq_documentation_block:
LOCUS       AC005149    98942 bp    DNA             INV       20-JUN-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS03364 (D237) and
            DS02911 (D145)), complete sequence.
ACCESSION   AC005149 AC004339 AC004278 AC003204 AC003205 AC002788 AC002789
            AC002790 AC002791 AC002792 AC002793 AC002794 AC002795 AC002796
            AC003206 AC002797 AC002798 AC002799 AC003207 AC002800 AC002801
            AC002802 AC003208
VERSION     AC005149.1  GI:3242679
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clones DS03364
            (D237) and DS02911 (D145)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 98942)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 24A1-24C5
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 98942)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
```

```
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
            to drosophila@mhgc.lbl.gov.
            Library locations: 4_36, 31_31.
FEATURES              Location/Qualifiers
     source           1. .98942
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /chromosome="2L"
                      /map="24A1-24C5"
                      /clone="Pls DS03364 (D237) and DS02911 (D145)"
                      /note="DS03364 (D237) extends from bp 1 to bp 79652 and
                      DS02911 (D145) extends from bp 15716 to bp 98943"
BASE COUNT    29273 a  20900 c  20166 g  28603 t
ORIGIN


alignment_scores:
            Quality:  131.00            Length:     56
              Ratio:  3.639               Gaps:      0
 Percent Similarity:  64.286   Percent Identity: 41.071

alignment_block:
 US-09-215-435-177_COPY_25_99 x AC005149/rev   ..

 Align seg 1/1  to reverse of: AC005149  from: 1  to: 98942

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         :::   |||          :::   |||   |||:::   :::   ::
    2464 GAGGACTGTCATCAGCCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTA 2415

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         :||||||   ||||||     ::::::   ||||||   |||||||:::
    2414 TCGCTACGCCTACAACGTGGATACACAATCCTGCGAGGAGTTCGTTTACG 2365

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::|||||   ||||||   ||||||||||:::   |||   :::   |||
    2364 GTGGATGTGCCGGCAACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAA 2315

      65 ValAlaCysValAlaLys 70
            ||||||:::::::|||
    2314 CAGGCATGTTTGGTAAAG 2297

 seq_name: gb_in1:AE003579

 seq_documentation_block:
 LOCUS       AE003579   304383 bp    DNA             INV       04-OCT-2000
 DEFINITION  Drosophila melanogaster genomic scaffold 142000013386046 section 12
             of 16, complete sequence.
 ACCESSION   AE003579 AE002638
 VERSION     AE003579.2  GI:10727349
 KEYWORDS    HTG.
 SOURCE      fruit fly.
   ORGANISM  Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE   1  (bases 1 to 304383)
   AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
             Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
             George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
             Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
             Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
             Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
             Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
             Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
             Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
             Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
             Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
             Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
             Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
             Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
             Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
             Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
             Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
             Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
             Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
             Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
             Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
             Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
             Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
             Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
             McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
             Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
             Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
             Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
             Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
             Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
             Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
             Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
             Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
             Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
             Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
             Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
             Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
             Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
   TITLE     The genome sequence of Drosophila melanogaster
   JOURNAL   Science 287 (5461), 2185-2195 (2000)
   MEDLINE   20196006
 REFERENCE   2  (bases 1 to 304383)
   AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
   TITLE     Direct Submission
   JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
             Rockville, MD, USA
 COMMENT     On Oct 9, 2000 this sequence version replaced gi:7295765.
 FEATURES              Location/Qualifiers
      source           1. .304383
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /chromosome="2L"
      mRNA             join(14340. .14462,14669. .15409,15464. .>15868)
                       /gene="CG2772"
                       /product="CT9119"
                       /db_xref="FLYBASE:FBan0002772"
                       /db_xref="FLYBASE:FBgn0031533"
      gene             <14340. .>15868
                       /gene="CG2772"
                       /map="23F1-23F3"
                       /db_xref="FLYBASE:FBan0002772"
                       /db_xref="FLYBASE:FBgn0031533"
      CDS              join(14358. .14462,14669. .15409,15464. .15868)
                       /gene="CG2772"
                       /note="CG2772 gene product"
                       /codon_start=1
                       /db_xref="FLYBASE:FBan0002772"
                       /db_xref="FLYBASE:FBgn0031533"
                       /protein_id="AAF51110.1"
                       /db_xref="GI:7295808"
                       /translation="MLAKVLTILAIVASAGAADDFDPFIDIPFKRLKTSAERIAEHGY
                       PAESHFVETPDGYVLNVFRIPHSPKLNSNGNEGESEASRPVVLIMHGLFSCSDCFLLN
                       GPEDALPYNYADAGYDVWLGNARGNIYSRNNTRLNVKHPYFWKFSWHEIGSIDLPATI
                       DYILAETGQQSLHYVGHSQGCTSFFVMGSYRPEYNAKIKTAHMLAPPVYMGNSTEGLI
```

```
                VSTAPLFGHHGIGSTLLENQVLLPQNAFIQRVLDTTCSNQPIMLSYCKTLAILWGGPE
                IGNLNQTLLPQIAETHPAGVSSNQAIHYIQSYASNDFRLYDWGSKRNLEYYGVSEPPA
                YDLTKITSELYLYYGLADGSANKQDISRLPDLLPNLALLHEVPDSTWGHLDFIFATEV
                KRVINDLVLDYSKAYDARETNNEI"
mRNA            join(18041. .18296,18504. .18607,18677. .19562,19620. .19772,
                19836. .>19886)
                /gene="CG2774"
                /product="CT9125"
                /db_xref="FLYBASE:FBan0002774"
                /db_xref="FLYBASE:FBgn0031534"
gene            <18041. .>19886
                /gene="CG2774"
                /map="23F1-23F3"
                /db_xref="FLYBASE:FBan0002774"
                /db_xref="FLYBASE:FBgn0031534"
CDS             join(18114. .18296,18504. .18607,18677. .19562,19620. .19772,
                19836. .19886)
                /gene="CG2774"
                /note="CG2774 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0002774"
                /db_xref="FLYBASE:FBgn0031534"
                /protein_id="AAF51109.1"
                /db_xref="GI:7295807"
                /translation="MEVESPEHTRDFAEVDINNGAATGSEDEEEEVPAPGSVTLDRNE
                SDLFVSALSPSSIGDIHPLQEVLTDDGDYFISIVVSDPQKIGDGMGSYLAYKVTTKTN
                IPKFKRSEFSTLRRFSDFLGIHDLLVGKYMRLGRIIPPAPSKNIIGSTKVKISPQQSE
                PGTPMTQEWVEIRRAALERFVHRTAQHPVLRVDLDFMNFLESDQELPRSVNTSALSGA
                GVIRLFNKVGETVNKITYKMDENDPWFDDKITEVESLDANLQKLHNAMKSLVTSRREL
                SLLTGLVAKSAAMLSTCEEHTGLSRALSNLADVEEKIELLRSEQANSDFFILAEFIKD
                YLGLFGAIKCIFHERVKAFQNWQYAQMQLSKRRENRGRFELANRADKLDQAQQEVDEW
                QGKVQRCQQQFDDISAEIKREMERFELTRVKDFKVNIIKYIEDQMAHQQQIVSYWEAF
                APFAREIV"
mRNA            join(20807. .20990,21059. .21597,21748. .>21915)
                /gene="CG12795"
                /product="CT38238"
                /db_xref="FLYBASE:FBan0012795"
                /db_xref="FLYBASE:FBgn0031535"
gene            <20807. .>21915
                /gene="CG12795"
                /map="23F1-23F3"
                /db_xref="FLYBASE:FBan0012795"
                /db_xref="FLYBASE:FBgn0031535"
CDS             join(20936. .20990,21059. .21597,21748. .21915)
                /gene="CG12795"
                /note="CG12795 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0012795"
                /db_xref="FLYBASE:FBgn0031535"
                /protein_id="AAF51108.1"
                /db_xref="GI:7295806"
                /translation="MEFPHLGQHCSEATCNRLDFLPVKCDSCDKVFCASHYSYDRHSC
                PGAYKKNVQVPVCPLCREPVPTPPGVEPDVTVGQHIDQQCKSESKKIYTNRCNAKGCK
                RKELIPVTCSQCRLNFCLRHRHTSDHDCKPVPASSTTSSRGGFQSIFKTSSDSRSMAA
                QAAERRQNRRPASNSISNTNSRPRPVQATQVQN1QGNMSEDEALARALALSIMEQDDA
                AESNRQAPAPSNAQQVAVGGGGNQQGNGKDKCLLS"
mRNA            complement(join(<22242. .22619,22678. .23381,23475. .23604,
                23663. .23823,23883. .24022,24076. .24563,24621. .25106,
                25190. .>25354))
                /gene="CG3248"
                /product="CT10923"
                /db_xref="FLYBASE:FBan0003248"
                /db_xref="FLYBASE:FBgn0031536"
gene            complement(<22242. .>25354)
                /gene="CG3248"
                /map="23F1-23F3"
                /db_xref="FLYBASE:FBan0003248"
                /db_xref="FLYBASE:FBgn0031536"
CDS             complement(join(22242. .22619,22678. .23381,23475. .23604,
                23663. .23823,23883. .24022,24076. .24563,24621. .25106,
                25190. .25354))
                /gene="CG3248"
                /note="CG3248 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0003248"
                /db_xref="FLYBASE:FBgn0031536"
                /protein_id="AAF51107.1"
                /db_xref="GI:7295805"
                /translation="MDDVERIQSENENLRKIRNRLMQWESKTDPLAALSIQQEEHLDV
                LTNLWRDSGPSVGPTDSSAAATSQSGDDIRLPAEGLQNTNEFLLWFADVSAEIEQRGD
                ADYHKYLQQLEQRKAECSHMLDQIAGAMERLGALCDEYDFVSQKTSALNTASEQLIEE
                QERLQELSHEIQRRLHYFSQVELLNQRLQSPTLSVASEAFRECLNKIDECLNYIEENP
                KFKDAAAYNVKYRQCLAKASGLVRNYVTSVINQATEATLHPKNNMPDASAALKAPDAA
                FALYYGKYQTAAAKVKRVAQLIESRSEHSLDYAQLMADLQQHYLAQRASVMSPAVNLS
                IQNVKVAHKGDHCSLTRSACAFLVHVCQDEQRLFYQFFSTGAPHLTVYLEGLCTILYD
                TMRPFIIHINHLETLAEICSILRIEMLEEHVQQNPVALEAFATIAHQLLQDVQERLVF
                RAHLYLQSDIQNFNPSSGDLAYPEKLEMMESIALSLQEPAPLRRSDSRNSMVSSVSSA
                VETESVATAYTVKQMIRRTLVCLSRLYRCVDRPIFQGLSQEALKLCIQSVSHAAGKIS
                ANKTPIDGELFEIKHLLILREQIAPFRVDFTVKETSLDFSKVKTAAFGLLQKRKQLFS
                MGSNNALLEFLLEGTPQIKEHLLDSRKEVDRQLKSVCEKYIKDAVQMLVGPLITFLEK
                AQSLLAQSTPATPQSPESTKASYVLRQSPWASPQQISSIIQETQRLIKAKLAVLQRSM
                QLYLSNRDTEFIIFRPIRNNIIQSFVKLEQLLTTNGYSTDDMIITSCPSAEQVSILLS
                SASILAAEGVASFAAAARKISTSSSVEGSTVGRKLSAMSNKSELVEEPKVEEVAGQIE
                VVPAEVQPPAKIEEEEVTETQTQANSE"
mRNA            join(25648. .25833,25904. .26518,26573. .28394,28452. .28603)
                /gene="CG8843"
                /product="CT9189"
                /db_xref="FLYBASE:FBan0008843"
                /db_xref="FLYBASE:FBgn0031537"
gene            <25648. .>28603
                /gene="CG8843"
                /map="23F1-23F3"
                /db_xref="FLYBASE:FBan0008843"
                /db_xref="FLYBASE:FBgn0031537"
CDS             join(25725. .25833,25904. .26518,26573. .28394,28452. .28590)
                /gene="CG8843"
                /note="CG8843 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0008843"
                /db_xref="FLYBASE:FBgn0031537"
                /protein_id="AAF51106.1"
                /db_xref="GI:7295804"
                /translation="MAPQPVVTGLSPKEGPPGTRVIIRGEFLGTRVQDLIGLKICGSD
                CLLSAEWKSPNKIIARTGPAKGKGDIIVTTLSGGVGTSTVQFRAYHETIGPLKESAVW

alignment_scores:
            Quality:  131.00            Length:     56
              Ratio:  3.639               Gaps:      0
 Percent Similarity:  64.286   Percent Identity: 41.071

alignment_block:
 US-09-215-435-177_COPY_25_99 x AE003579/rev   ..

 Align seg 1/1  to reverse of: AE003579  from: 1  to: 304383

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         :::   |||          :::   |||   |||:::   :::   ::
  265613 GAGGACTGTCATCAGCCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTA 265564

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         :||||||   ||||||      ::::::   ||||||   |||||||::::
  265563 TCGCTACGCCTACAACGTGGATACACAATCCTGCGAGGAGTTCGTTTACG 265514

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||   |||||||||:::   |||   :::   |||
  265513 GTGGATGTGCCGGCAACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAA 265464

      65 ValAlaCysValAlaLys 70
            ||||||::::::|||
  265463 CAGGCATGTTTGGTAAAG 265446

seq_name: gb_om:MEELACP

seq_documentation_block:
LOCUS       MEELACP       419 bp    mRNA            MAM       15-APR-1998
DEFINITION  Macropus eugenii mRNA for early lactation protein (ELP).
```

```
ACCESSION   AJ000490
VERSION     AJ000490.1  GI:3059123
KEYWORDS    early lactation protein; ELP gene.
SOURCE      tammar wallaby.
  ORGANISM  Macropus eugenii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Simpson,K.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1997) Simpson K.J., Division of Molecular Biology
            and Genetics, Victorian Institute of Animal Science, 475 Mickleham
            Rd, Attwood, Victoria, 3049, AUSTRALIA
  REMARK    Revised by [3]
REFERENCE   2  (bases 1 to 419)
  AUTHORS   Simpson,K.J., Shaw,D. and Nicholas,K.R.
  TITLE     Asynchronous expression of a putative protease inhibitor gene in
            the mammary gland during lactation in the Tammar wallaby, Macropus
            eugenii
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 419)
  AUTHORS   Simpson,K.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1998) Simpson K.J., Division of Molecular Biology
            and Genetics, Victorian Institute of Animal Science, 475 Mickleham
            Rd, Attwood, Victoria, 3049, AUSTRALIA
COMMENT     On Apr 18, 1998 this sequence version replaced gi:3046757.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Macropus eugenii"
                     /db_xref="taxon:9315"
     gene            1. .252
                     /gene="ELP"
     CDS             <1. .252
                     /gene="ELP"
                     /codon_start=1
                     /evidence=experimental
                     /product="early lactation protein"
                     /protein_id="CAA04128.1"
                     /db_xref="GI:3046758"
                     /db_xref="SPTREMBL:O62845"
                     /translation="EKCLDQIQVNSLENLSLLVPSLCLLPPVRGNCSSQILHYFYNTT
                     SRTCETFIYSGCNGNRNNFNSEEYCLKTCRRNKNRNNNN"
     source          1. .255
                     /organism="Macropus eugenii"
                     /db_xref="taxon:9315"
                     /tissue_type="lactating mammary gland"
     polyA_signal    370. .375
     polyA_site      386
BASE COUNT      139 a     98 c     72 g    110 t
ORIGIN


alignment_scores:
            Quality:  130.00            Length:     41
              Ratio:  4.333               Gaps:      0
 Percent Similarity:  73.171   Percent Identity: 56.098

alignment_block:
 US-09-215-435-177_COPY_25_99 x MEELACP   ..

 Align seg 1/1  to: MEELACP  from: 1  to: 419

      17 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 33
         |||   |||           |||:::|||   :::       :::::::||
      67 TGTCTGCTCCCACCTGTGAGGGGTAACTGCAGTTCTCAAATTCTTCACTA 116

      33 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 50
         ||||||||||   ||||||:::   |||||||||||:::::::|||||||
     117 TTTCTACAACACAACCTCTCGAACCTGTGAGACTTTCATCTACAGTGGCT 166

      50 ysAsnGlyAsnLeuAsnAsnPhe 57
         ||||||||||   |||||||||
```

```
     167 GCAATGGCAATAGGAACAATTTT 189

seq_name: gb_pat2:I63559

seq_documentation_block:
LOCUS       I63559       1455 bp    DNA              PAT       07-OCT-1997
DEFINITION  Sequence 6 from patent US 5663143.
ACCESSION   I63559
VERSION     I63559.1  GI:2481132
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1455)
  AUTHORS   Ley,A.Charles, Ladner,R.Charles, Guterman,S.Kosow,
            Roberts,B.Lindsay, Markland,W. and Kent,R.Baribault.
  TITLE     Engineered human-derived kunitz domains that inhibit human
            neutrophil elastase
  JOURNAL   Patent: US 5663143-A 6 02-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .1455
                     /organism="unknown"
BASE COUNT      337 a    316 c    366 g    436 t
ORIGIN


alignment_scores:
            Quality:  130.00            Length:     72
              Ratio:  3.333               Gaps:      2
 Percent Similarity:  54.167   Percent Identity: 41.667

alignment_block:
 US-09-215-435-177_COPY_25_99 x I63559   ..

 Align seg 1/1  to: I63559  from: 1  to: 1455

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||      :::|||   |||:::|||      :::   |||   |||
      49 GGCGCCAAAGAAGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCAT 98

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
            :::      ||||||||||||||||   ||||||      |||||||
      99 GGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGA 148

      44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
         |||||   ::::::|||||   ||||||   |||||||||      :::
     149 CTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAA 198

      61 IleGlu.............ArgGluVal..................... 65
            |||               |||   |||
     199 AAGGAGTGTCTGCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGA 248

      66 .AlaCysValAlaLys 70
         :::|||:::||||||
     249 AAGTTGTTTAGCAAAA 264

seq_name: gb_pr9:HSAIMBG3

seq_documentation_block:
LOCUS       HSAIMBG3     4627 bp    DNA              PRI       06-OCT-1999
DEFINITION  Human gene for alpha-1-microglobulin-bikunin, exons 7-10 (encoding
            linker peptide and bikunin).
ACCESSION   X54818
VERSION     X54818.1  GI:28575
KEYWORDS    alpha-1-microglobulin; bikunin; glycoprotein; proteinase inhibitor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4627)
  AUTHORS   Vetr,H. and Gebhard,W.
  TITLE     Structure of the human alpha 1-microglobulin-bikunin gene
  JOURNAL   Biol. Chem. Hoppe-Seyler 371 (12), 1185-1196 (1990)
```

```
  MEDLINE   91214554
REFERENCE   2  (bases 1 to 4627)
  AUTHORS   Gebhard,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1990) Gebhard W., Universitaet Muenchen, Klinikum
            Grosshadern, HNO Forschungslabor, Marchioninistr 15, 8000 Muenchen
            70. Germany
COMMENT     Sequence is approximately 6 kbp downstream of x54817.
FEATURES             Location/Qualifiers
     source          1. .4627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="leukocyte"
                     /clone_lib="EMBL3 vector"
     prim_transcript <1. .2801
                     /gene="alpha-1-microglobulin-bikunin"
     intron          <1. .171
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=6
     gene            1. .2801
                     /gene="alpha-1-microglobulin-bikunin"
     exon            172. .253
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=7
                     /product="linker peptide and bikunin"
     mRNA            join(172. .253,1339. .1506,1832. .2005,2658. .2801)
                     /partial
                     /gene="alpha-1-microglobulin-bikunin"
                     /usedin=X54816:A1MB_mRNA_a
                     /usedin=X54816:A1MB_mRNA_b
     CDS             join(172. .253,1339. .1506,1832. .2005,2658. .2689)
                     /partial
                     /gene="alpha-1-microglobulin-bikunin"
                     /note="alpha-1-microglobulin"
                     /codon_start=1
                     /usedin=X54816:A1MB_cds
                     /product="bikunin"
                     /protein_id="CAA38587.1"
                     /db_xref="GI:579676"
                     /db_xref="SPTREMBL:P78491"
                     /translation="RVRRAVLPQEEEGSGGGQLVTEVTKKEDSCQLGYSAGPCMGMTS
                     RYFYNGTSMACETFQYGGCMGNGNNFVTEKECLQTCRTVAACNLPIVRGPCRAFIQLW
                     AFDAVKGKCVLFPYGGCQGNGNKFYSEKECREYCGVPGDGDEELLRFSN"
     intron          254. .1338
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=7
     exon            1339. .1506
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=8
                     /product="bikunin"
     intron          1507. .1831
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=8
     exon            1832. .2005
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=9
                     /product="bikunin"
     intron          2006. .2657
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=9
     exon            2658. .2801
                     /gene="alpha-1-microglobulin-bikunin"
                     /number=10
                     /product="bikunin"
     polyA_signal    2773. .2778
                     /gene="alpha-1-microglobulin-bikunin"
     polyA_site      2801
                     /gene="alpha-1-microglobulin-bikunin"
     repeat_unit     4288. .4612
                     /rpt_family="Alu"
BASE COUNT     1088 a   1133 c   1292 g   1108 t      6 others
ORIGIN
```

```
alignment_scores:
           Quality:  129.50              Length:     59
             Ratio:  3.597                 Gaps:      1
 Percent Similarity:  61.017   Percent Identity: 45.763

alignment_block:
 US-09-215-435-177_COPY_25_99 x HSAIMBG3   ..

 Align seg 1/1  to: HSAIMBG3  from: 1  to: 4627

       9 IleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySe 25
         :::|||   :::|||   |||   |||:::|||      :::   |||
    1323 GTCTGCCCCTCCCTA...GATTCCTGCCAGCTGGGCTACTCGGCCGGTCC 1369

      25 rCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgC 42
          |||      :::      ||||||||||||||   ||||||      |
    1370 CTGCATGGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCT 1419

      42 ysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLys 58
         ||||||||||   ::::::||||||   ||||||   |||||||||
    1420 GTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTC 1469

      59 LeuLysIleGluArgGluValAlaCys 67
         :::   |||         :::|||
    1470 ACAGAAAAGGAGTGTCTGCAGACCTGC 1496

seq_name: gb_pr10:HUMITILC08

seq_documentation_block:
LOCUS       HUMITILC08   5423 bp    DNA             PRI       06-JAN-1995
DEFINITION  Human inter-alpha-trypsin inhibitor light chain (ITI) gene, exons
            7-10.
ACCESSION   M88249 M27417
VERSION     M88249.1  GI:186598
KEYWORDS    inter-alpha-trypsin inhibitor light chain.
SEGMENT     8 of 8
SOURCE      Homo sapiens (tissue library: of T.Maniatis and Clontech) fetus
            liver DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5423)
  AUTHORS   Diarra-Mehrpour,M., Bourguignon,J., Sesboue,R., Salier,J.P.,
            Leveillard,T. and Martin,J.P.
  TITLE     Structural analysis of the human inter-alpha-trypsin inhibitor
            light-chain gene
  JOURNAL   Eur. J. Biochem. 191 (1), 131-139 (1990)
  MEDLINE   90336621
FEATURES             Location/Qualifiers
     source          1. .5423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="fetus"
                     /tissue_type="liver"
                     /tissue_lib="of T.Maniatis and Clontech"
                     /map="9q32-q33"
     gene            join(M88165.1:532. .1176,M88243.1:1. .489,M88244.1:1. .618,
                     M88245.1:1. .340,M88246.1:1. .690,M88247.1:1. .335,1. .3604)
                     /gene="ITI light chain"
     mRNA            join(M88165.1:690. .900,M88243.1:265. .407,
                     M88244.1:323. .399,M88245.1:198. .314,M88246.1:138. .239,
                     M88247.1:192. .238,976. .1057,2145. .2312,2638. .2811,
                     3464. .3604)
                     /gene="ITI light chain"
                     /product="inter-alpha-trypsin inhibitor light chain"
     gene            join(M88165.1:690. .1176,M88243.1:1. .489,M88244.1:1. .618,
                     M88245.1:1. .340,M88246.1:1. .690,M88247.1:1. .335,
                     M88248.1:1. .287,1. .3495)
                     /gene="ITIL"
     mRNA            join(M88165.1:694. .900,M88243.1:265. .407,
                     M88244.1:323. .399,M88245.1:198. .314,M88246.1:138. .239,
                     M88247.1:192. .238,976. .1057,2145. .2312,2638. .2811,
```

```
                3464. .3604)
                /gene="ITI light chain"
                /product="inter-alpha-trypsin inhibitor light chain"
     CDS        join(M88165.1:784. .900,M88243.1:265. .407,
                M88244.1:323. .399,M88245.1:198. .314,M88246.1:138. .239,
                M88247.1:192. .238,976. .1057,2145. .2312,2638. .2811,
                3464. .3495)
                /gene="ITIL"
                /codon_start=1
                /db_xref="GDB:G00-120-696"
                /product="inter-alpha-trypsin inhibitor light chain"
                /protein_id="AAA59196.1"
                /db_xref="GI:186600"
                /translation="MRSLGALLLLLSACLAVSAGPVPTPPDNIQVQENFNISRIYGKW
                YNLAIGSTCPWLKKIMDRMTVSTLVLGEGATEAEISMTSTRWRKGVCEETSGAYEKTD
                TDGKFLYHKSKWNITMESYVVHTNYDEYAIFLTKKFSRHHGPTITAKLYGRAPQLRET
                LLQDFRVVAQGVGIPEDSIFTMADRGECVPGEQEPEPILIPRVRRAVLPQEEEGSGGG
                QLVTEVTKKEDSCQLGYSAGPCMGMTSRYFYNGTSMACETFQYGGCMGNGNNFVTEKE
                CLQTCRTVAACNLPIVRGPCRAFIQLWAFDAVKGKCVLFPYGGCQGNGNKFYSEKECR
                EYCGVPGDGDEELLRFSN"
     intron     order(M88247.1:239. .335,M88248.1:1. .287,1. .975)
                /gene="ITI light chain"
                /note="intron is about 5.8 kb."
                /number=6
     TATA_signal 900. .906
                /gene="ITI light chain"
     exon       976. .1057
                /gene="ITI light chain"
                /number=7
     intron     1058. .2144
                /gene="ITI light chain"
                /number=7
     exon       2145. .2312
                /gene="ITI light chain"
                /number=8
     intron     2313. .2637
                /gene="ITI light chain"
                /number=8
     exon       2638. .2811
                /gene="ITI light chain"
                /number=9
     intron     2812. .3463
                /gene="ITI light chain"
                /number=9
     exon       3464. .3604
                /gene="ITI light chain"
                /number=10
     repeat_unit complement(5112. .5402)
                /rpt_family="Alu repetitive element"
BASE COUNT     1258 a   1333 c   1450 g   1382 t
ORIGIN


alignment_scores:
            Quality:  129.50              Length:     59
              Ratio:  3.597                 Gaps:      1
 Percent Similarity:  61.017    Percent Identity: 45.763

alignment_block:
 US-09-215-435-177_COPY_25_99 x HUMITILC08   ..

 Align seg 1/1  to: HUMITILC08  from: 1  to: 5423

       9 IleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySe 25
         :::|||   :::|||   |||   |||:::|||      :::   |||
    2129 GTCTGCCCCTCCCTA...GATTCCTGCCAGCTGGGCTACTCGGCCGGTCC 2175

      25 rCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgC 42
          |||      :::      ||||||||||||||||   ||||||      |
    2176 CTGCATGGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCT 2225

      42 ysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLys 58
       . ||||||||||   ::::::||||||   ||||||   |||||||||
          ' %
    2226 GTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTC 2275

      59 LeuLysIleGluArgGluValAlaCys 67
         :::   |||         :::|||
    2276 ACAGAAAAGGAGTGTCTGCAGACCTGC 2302

seq_name: gb_htg19:AL137850

seq_documentation_block:
LOCUS       AL137850   193939 bp    DNA              HTG       12-APR-2001
DEFINITION  Homo sapiens chromosome 9 clone RP11-56P10 map q31.3-33.3, ***
            SEQUENCING IN PROGRESS ***, 3 unordered pieces.
ACCESSION   AL137850
VERSION     AL137850.9  GI:13624463
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193939)
  AUTHORS   Corby,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 14, 2001 this sequence version replaced gi:13160162.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA56P10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 193037 bases at least Q40
            Consensus quality: 193307 bases at least Q30
            Consensus quality: 193499 bases at least Q20
            Insert size: 193739; sum-of-contigs
            Insert size: 175330; 1.5% error; agarose-fp
            Quality coverage: 9.36x in Q20 bases; sum-of-contigs Quality
            coverage: 10.65x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    92351: contig of 92351 bp in length
            *    92352 92451: gap of      100 bp
            *    92452   111748: contig of 19297 bp in length
            *   111749 111848: gap of      100 bp
            *   111849   193939: contig of 82091 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193939
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q31.3-33.3"
                     /clone="RP11-56P10"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .92351
                     /note="assembly_fragment:01539
                     fragment_chain:1"
     misc_feature    92452. .111748
                     /note="assembly_fragment:01080
                     fragment_chain:1"
     misc_feature    111849. .193939
```

```
                     /note="assembly_fragment:01170
                     fragment_chain:1"
BASE COUNT    48491 a  46496 c  47840 g  50912 t    200 others
ORIGIN


alignment_scores:
            Quality:  129.50             Length:     59
              Ratio:  3.597               Gaps:      1
 Percent Similarity:  61.017   Percent Identity: 45.763

alignment_block:
 US-09-215-435-177_COPY_25_99 x AL137850/rev   ..

 Align seg 1/1  to reverse of: AL137850  from: 1  to: 193939

       9 IleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySe 25
         :::|||   :::|||   |||   |||:::|||      :::   |||
  103652 GTCTGCCCCTCCCTA...GATTCCTGCCAGCTGGGCTACTCGGCCGGTCC 103606

      25 rCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgC 42
          |||    :::      ||||||||||||||   ||||||      |
  103605 CTGCATGGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCT 103556

      42 ysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLys 58
         ||||||||||   ::::::||||||   ||||||   |||||||||
  103555 GTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTC 103506

      59 LeuLysIleGluArgGluValAlaCys 67
          :::   |||         :::|||
  103505 ACAGAAAAGGAGTGTCTGCAGACCTGC 103479

seq_name: gb_htg3:AC011161

seq_documentation_block:
LOCUS       AC011161   200532 bp    DNA              HTG       30-MAR-2000
DEFINITION  Homo sapiens clone RP11-9L6, WORKING DRAFT SEQUENCE, 33 unordered
            pieces.
ACCESSION   AC011161
VERSION     AC011161.3  GI:7341865
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-9L6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6139126.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3019
                Center clone name: 9_L_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 175452 bases at least Q40
                Consensus quality: 186874 bases at least Q30
                Consensus quality: 191961 bases at least Q20
                Insert size: 200000; agarose-fp
                Insert size: 197332; sum-of-contigs
                Quality coverage: 3.2 in Q20 bases; agarose-fp
                Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1310: contig of 1310 bp in length
            *     1311 1410: gap of      100 bp
            *     1411     3606: contig of 2196 bp in length
            *     3607 3706: gap of      100 bp
            *     3707     5209: contig of 1503 bp in length
            *     5210 5309: gap of      100 bp
            *     5310     7094: contig of 1785 bp in length
            *     7095 7194: gap of      100 bp
            *     7195     9372: contig of 2178 bp in length
            *     9373 9472: gap of      100 bp
            *     9473    11334: contig of 1862 bp in length
            *    11335 11434: gap of      100 bp
            *    11435    13898: contig of 2464 bp in length
            *    13899 13998: gap of      100 bp
            *    13999    15543: contig of 1545 bp in length
            *    15544 15643: gap of      100 bp
            *    15644    17721: contig of 2078 bp in length
            *    17722 17821: gap of      100 bp
            *    17822    20076: contig of 2255 bp in length
            *    20077 20176: gap of      100 bp
            *    20177    22360: contig of 2184 bp in length
            *    22361 22460: gap of      100 bp
            *    22461    25056: contig of 2596 bp in length
            *    25057 25156: gap of      100 bp
            *    25157    29276: contig of 4120 bp in length
            *    29277 29376: gap of      100 bp
            *    29377    33914: contig of 4538 bp in length
            *    33915 34014: gap of      100 bp
            *    34015    39784: contig of 5770 bp in length
            *    39785 39884: gap of      100 bp
            *    39885    43727: contig of 3843 bp in length
            *    43728 43827: gap of      100 bp
            *    43828    48356: contig of 4529 bp in length
            *    48357 48456: gap of      100 bp
            *    48457    53189: contig of 4733 bp in length
            *    53190 53289: gap of      100 bp
            *    53290    58926: contig of 5637 bp in length
            *    58927 59026: gap of      100 bp
            *    59027    64871: contig of 5845 bp in length
            *    64872 64971: gap of      100 bp
            *    64972    70550: contig of 5579 bp in length
            *    70551 70650: gap of      100 bp
            *    70651    76382: contig of 5732 bp in length
            *    76383 76482: gap of      100 bp
            *    76483    82846: contig of 6364 bp in length
            *    82847 82946: gap of      100 bp
            *    82947    91987: contig of 9041 bp in length
```

```
            *    91988 92087: gap of      100 bp
            *    92088    98742: contig of 6655 bp in length
            *    98743 98842: gap of      100 bp
            *    98843   105574: contig of 6732 bp in length
            *   105575 105674: gap of      100 bp
            *   105675   114985: contig of 9311 bp in length
            *   114986 115085: gap of      100 bp
            *   115086   124472: contig of 9387 bp in length
            *   124473 124572: gap of      100 bp
            *   124573   134778: contig of 10206 bp in length
            *   134779 134878: gap of      100 bp
            *   134879   150202: contig of 15324 bp in length
            *   150203 150302: gap of      100 bp
            *   150303   166267: contig of 15965 bp in length
            *   166268 166367: gap of      100 bp
            *   166368   183386: contig of 17019 bp in length
            *   183387 183486: gap of      100 bp
            *   183487   200532: contig of 17046 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200532
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-9L6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1310
                     /note="assembly_fragment"
     misc_feature    1411. .3606
                     /note="assembly_fragment"
     misc_feature    3707. .5209
                     /note="assembly_fragment"
     misc_feature    5310. .7094
                     /note="assembly_fragment"
     misc_feature    7195. .9372
                     /note="assembly_fragment"
     misc_feature    9473. .11334
                     /note="assembly_fragment"
     misc_feature    11435. .13898
                     /note="assembly_fragment"
     misc_feature    13999. .15543
                     /note="assembly_fragment"
     misc_feature    15644. .17721
                     /note="assembly_fragment"
     misc_feature    17822. .20076
                     /note="assembly_fragment"
     misc_feature    20177. .22360
                     /note="assembly_fragment"
     misc_feature    22461. .25056
                     /note="assembly_fragment"
     misc_feature    25157. .29276
                     /note="assembly_fragment"
     misc_feature    29377. .33914
                     /note="assembly_fragment"
     misc_feature    34015. .39784
                     /note="assembly_fragment"
     misc_feature    39885. .43727
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    43828. .48356
                     /note="assembly_fragment"
     misc_feature    48457. .53189
                     /note="assembly_fragment"
     misc_feature    53290. .58926
                     /note="assembly_fragment"
     misc_feature    59027. .64871
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    64972. .70550
                     /note="assembly_fragment"
     misc_feature    70651. .76382
                     /note="assembly_fragment"
     misc_feature    76483. .82846
                     /note="assembly_fragment"
     misc_feature    82947. .91987
                     /note="assembly_fragment"
     misc_feature    92088. .98742
                     /note="assembly_fragment"
     misc_feature    98843. .105574
                     /note="assembly_fragment"
     misc_feature    105675. .114985
                     /note="assembly_fragment"
     misc_feature    115086. .124472
                     /note="assembly_fragment"
     misc_feature    124573. .134778
                     /note="assembly_fragment"
     misc_feature    134879. .150202
                     /note="assembly_fragment"
     misc_feature    150303. .166267

alignment_scores:
            Quality:  129.50              Length:     59
              Ratio:  3.597                 Gaps:      1
 Percent Similarity:  61.017   Percent Identity: 45.763

alignment_block:
 US-09-215-435-177_COPY_25_99 x AC011161/rev   ..

 Align seg 1/1  to reverse of: AC011161  from: 1  to: 200532

       9 IleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySe 25
         :::|||   :::|||   |||   |||:::|||      :::   |||
  192488 GTCTGCCCCTCCCTA...GATTCCTGCCAGCTGGGCTACTCGGCCGGTCC 192442

      25 rCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgC 42
          |||      :::      ||||||||||||||||   ||||||      |
  192441 CTGCATGGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCT 192392

      42 ysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLys 58
         ||||||||||   ::::::||||||   ||||||   |||||||||
  192391 GTGAGACTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTC 192342

      59 LeuLysIleGluArgGluValAlaCys 67
           :::   |||         :::|||
  192341 ACAGAAAAGGAGTGTCTGCAGACCTGC 192315
```

OM of: US-09-215-435-177_COPY_25_99 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:28 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612	+	419.00	997.35	1.4e-47	526	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738	+	409.00	972.94	3.2e-46	539	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455	+	307.00	733.42	7.0e-33	267	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569	+	307.00	733.28	7.1e-33	271	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32013	+	133.00	262.21	1.2e-06	42521		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90070	+	133.00	262.21	1.2e-06	42521		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35164	+	130.00	304.97	5.1e-09	285	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV40046	+	128.50	296.98	1.4e-08	441	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN81432	+	128.50	286.72	5.3e-08	1232		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT78950	+	128.50	286.53	5.5e-08	1255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77939	+	128.50	285.13	6.5e-08	1444		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ06878	+	127.00	301.71	7.8e-09	191	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46936	+	127.00	293.89	2.1e-08	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79081	+	126.00	292.64	2.5e-08	372	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79088	+	126.00	290.37	3.3e-08	467	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT15870	+	125.00	295.38	1.8e-08	222	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79079	+	124.00	295.93	1.6e-08	165	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79080	+	124.00	287.80	4.6e-08	372	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79083	+	124.00	286.24	5.7e-08	435	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79086	+	124.00	285.53	6.2e-08	467	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79087	+	124.00	284.27	7.3e-08	530	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46935	+	123.00	284.22	7.3e-08	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79077	+	122.00	291.85	2.8e-08	153	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV86987	+	121.00	275.89	2.1e-07	593	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ90469	+	121.00	270.89	4.1e-07	979	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46934	+	120.00	276.97	1.9e-07	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ51001	+	120.00	268.77	5.3e-07	950	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46933	+	119.00	274.55	2.5e-07	418	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29017	+	118.00	280.55	1.2e-07	180	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29024	+	118.00	277.19	1.8e-07	252	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV33175	+	118.00	274.03	2.7e-07	346	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32238	+	118.00	262.34	1.2e-06	1114		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35169	+	117.00	274.90	2.4e-07	249	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT10945	+	117.00	261.99	1.3e-06	906	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ39947	+	117.00	258.95	1.9e-06	1228		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ10992	+	116.00	266.05	7.6e-07	474	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ10993	+	116.00	262.25	1.2e-06	693	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51316	+	116.00	259.51	1.7e-06	912	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51317	+	116.00	259.51	1.7e-06	912	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ81396	+	116.00	259.33	1.8e-06	928	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46644	+	116.00	259.15	1.8e-06	945		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46931	+	116.00	259.15	1.8e-06	945		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT16056	+	116.00	257.96	2.1e-06	106		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1989.DAT:AAN90108	+	116.00	255.01	3.1e-06	143		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32167	+	116.00	255.01	3.1e-06	143		

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97612

seq_documentation_block:
ID AAX97612 standard; DNA; 526 BP.
XX
AC AAX97612;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 76.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; AAY35928.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 210; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 526 BP; 162 A; 100 C; 99 G; 165 T; 0 other;

alignment_scores:
Quality: 419.00 Length: 75
Ratio: 5.587 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-177_COPY_25_99 x AAX97612 ..

Align seg 1/1 to: AAX97612 from: 1 to: 526

```
      1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     94 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 143

     17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    144 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 193

     34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    194 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCGGCTGT 243

     51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    244 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAGCCTG 293

     67 sValAlaLysTyrLysProProArg 75
        |||||||||||||||||||||||||
    294 TGTTGCAAAATACAAACCACCGAGG 318
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97738

seq_documentation_block:

```
ID   AAX97738 standard; DNA; 539 BP.
XX
AC   AAX97738;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 303.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY36054.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 376; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 539 BP; 165 A; 103 C; 97 G; 162 T; 12 other;
```

```
alignment_scores:
            Quality:  409.00              Length:     75
              Ratio:  5.453                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.333

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAX97738   ..

 Align seq 1/1  to: AAX97738  from: 1  to: 539

      1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    104 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 153

     17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    154 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 203

     34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||:::|||
    204 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 253

     51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
        ||||||||||||||||||||||||||||||||||||||||||||||:::||
    254 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAKCCTG 303

     67 sValAlaLysTyrLysProProArg 75
        |||||||||||||||||||||||||
    304 TGTTGCAAAATACAAACCACCGAGG 328
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00455

seq_documentation_block:

```
ID   AAC00455 standard; cDNA; 267 BP.
XX
AC   AAC00455;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 453.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00449.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 453; 71pp + CD-ROM; English.
XX
```

```
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 267 BP; 74 A; 57 C; 52 G; 84 T; 0 other;


alignment_scores:
            Quality:  307.00              Length:     55
              Ratio:  5.582                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.182

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAC00455   ..

 Align seg 1/1  to: AAC00455  from: 1  to: 267

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 152

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 202

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
     203 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 252

      51 AsnGlyAsnLeuAsn 55
         ||||||||||||||
     253 AATGGCAACCTTAAC 267

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51569

seq_documentation_block:
ID   AAX51569 standard; cDNA; 271 BP.
XX
AC   AAX51569;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:148.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12791.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 253-254; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 271 BP; 75 A; 57 C; 55 G; 84 T; 0 other;


alignment_scores:
            Quality:  307.00              Length:     55
              Ratio:  5.582                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.182

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAX51569   ..

 Align seg 1/1  to: AAX51569  from: 1  to: 271

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     104 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 153

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 203

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
     204 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 253

      51 AsnGlyAsnLeuAsn 55
         ||||||||||||||
     254 AATGGCAACCTTAAC 268

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ32013

seq_documentation_block:
ID   AAZ32013 standard; DNA; 42521 BP.
XX
AC   AAZ32013;
XX
DT   10-JAN-2000  (first entry)
XX
DE   Human METH1 related EST U64857.
XX
KW   Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW   cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW   angiogenesis inhibitor; abnormal wound healing; inflammation;
KW   rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
```

```
KW   diabetic retinopathy; macula degeneration; haemangioma; detection;
KW   arterial-venous malformation; immune deficiency; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937660-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-US01313.
XX
PR   23-JAN-1998;   98US-0072298.
PR   28-AUG-1998;   98US-0098539.
XX
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM;
XX
DR   WPI; 1999-590684/50.
XX
PT   New isolated metalloprotease thrombospondin polypeptides, useful for
PT   treating hyperproliferative disorders, cancers or autoimmune disorders
PT   -
XX
PS   Disclosure; Page 256-283; 457pp; English.
XX
CC   AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC   metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC   respectively. METH1 and METH2 have been found to be potent inhibitors of
CC   angiogenesis both in vitro and in vivo. They can be used for treating
CC   cancer and other disorders related to angiogenesis including abnormal
CC   wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC   endometrial bleeding disorders, diabetic retinopathy, some forms of
CC   macula degeneration, haemangiomas, and arterial-venous malformations.
CC   They may be useful in treating deficiencies or disorders of the immune
CC   system, by activating or inhibiting the proliferation, differentiation,
CC   or mobilisation (chemotaxis) of immune cells. The etiology of these
CC   immune deficiencies or disorders may be genetic, somatic, such as
CC   cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC   toxins), or infectious. They can also be used to treat inflammatory
CC   conditions, both chronic and acute conditions. The products can also be
CC   used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC   AAY49511 represent sequences given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 42521 BP; 13493 A; 8055 C; 8365 G; 12608 T; 0 other;


alignment_scores:
            Quality:  133.00              Length:     62
              Ratio:  3.500                Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAZ32013   ..

 Align seg 1/1  to: AAZ32013  from: 1  to: 42521

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   :::   ||||||   |||:::|||           |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |         |||      |||:::||||||   :::|||  :::||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      43 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 59
         ||||||||:::::::::||||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      60 LysIleGluArgGluValAlaCysValAlaLysTyr 71
         |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90070

seq_documentation_block:
ID   AAC90070 standard; DNA; 42521 BP.
XX
AC   AAC90070;
XX
DT   19-MAR-2001  (first entry)
XX
DE   U64857 cDNA clone.
XX
KW   METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW   cancer therapy; benign tumour; ocular angiogenic disease;
KW   rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW   vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW   scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW   coronary collateral; cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW   plaque neovascularisation; telangiectasia; haemophiliac joint; EST;
KW   angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW   Crohn's disease; atherosclerosis; birth control; ss.
XX
OS   Unidentified.
XX
PN   WO200071577-A1.
XX
PD   30-NOV-2000.
XX
PF   25-MAY-2000; 2000WO-US14462.
XX
PR   25-MAY-1999;   99US-0318208.
PR   20-JUL-1999;   99US-0144882.
PR   10-AUG-1999;   99US-0147823.
PR   13-AUG-1999;   99US-0373658.
PR   22-DEC-1999;   99US-0171503.
PR   22-FEB-2000; 2000US-0183792.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
PA   (JONA/) JONAK Z L.
PA   (TRUL/) TRULLI S H.
PA   (FORN/) FORNWALD J A.
PA   (TERR/) TERRETT J A.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM,  Jonak ZL,  Trulli SH;
PI   Fornwald JA,  Terrett JA;
XX
DR   WPI; 2001-025136/03.
XX
PT   METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT   inhibit angiogensis in the treatment of disorders such as cancer,
PT   rheumatoid arthritis and psoriasis -
XX
PS   Claim 7; Pages 556-583; 768pp; English.
XX
CC   The present invention relates to human METH1 and METH2, (ME for
CC   metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC   The present sequence is an expressed sequence tag (EST) for METH. METH
CC   can be used for inhibiting angiogenesis in an individual, and for
CC   treating cancer, benign tumours, an ocular angiogenic disease,
CC   rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC   vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC   scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC   coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC   ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
```

```
CC   fibromuscular dysplasia, wound granulation, Crohn's disease or
CC   atherosclerosis. METH can also be used in birth control. METH can also
CC   be used in diagnostic methods for the prognosis of cancer.
XX
SQ   Sequence 42521 BP; 13493 A; 8055 C; 8365 G; 12608 T; 0 other;


alignment_scores:
           Quality:  133.00              Length:     62
             Ratio:  3.500                 Gaps:      0
 Percent Similarity:  61.290   Percent Identity: 40.323

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAC90070   ..

 Align seg 1/1  to: AAC90070  from: 1  to: 42521

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   :::   ||||||   |||:::|||           |||   ||
    6370 TGTCAAAATCATAAGGATGCTTGTCAACTTCCAAAGGTTCAAGGACCATG 6419

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |         |||      |||:::||||||   :::|||   :::||||
    6420 CTCTGGAAAGCATTCCTATTATTACTACAACACTGCCAGTCATCAATGCG 6469

      43 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 59
         ||||||||:::::::::||||||   ||||||   |||:::|||
    6470 AGACGTTCACTTATGGTGGCTGCCTCGGAAATACTAACAGATTCGCTACC 6519

      60 LysIleGluArgGluValAlaCysValAlaLysTyr 71
            |||   ::::::   |||   :::|||:::
    6520 ATTGAGGAGTGTCAAGCGAGATGCCCGAGTAAGTTC 6555

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT35164

seq_documentation_block:
ID   AAT35164 standard; DNA; 285 BP.
XX
AC   AAT35164;
XX
DT   14-FEB-1997  (first entry)
XX
DE   ITI-D1 Kunitz domain fusion gene.
XX
KW   Aprotinin; Kunitz domain; human neutrophil elastase; hNE;
KW   connective tissue; alpha 1 protease inhibitor; API; neutrophil;
KW   alphal antitrypsin; respiratory disorder; cystic fibrosis;
KW   smokers emphysema; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..285
FT                   /*tag=  a
FT                   /product= IIIsp::itiDl::matureIII (initial fragment)
FT                            fusion protein
FT   misc_feature    1..3
FT                   /*tag=  b
FT                   /transl_except= GTG encodes Met
FT   sig_peptide     1..54
FT                   /*tag=  c
FT                   /note= "M13 III protein signal peptide"
FT   mat_peptide     55..234
FT                   /*tag=  d
FT                   /note= "ITI-D1 Kunitz domain"
FT   mat_peptide     235..285
FT                   /*tag=  e
FT                   /note= "M13 III protein"
XX
PN   WO9620278-A2.
XX
PD   04-JUL-1996.
XX
PF   15-DEC-1995;   95WO-US16349.
XX
PR   16-DEC-1994;   94US-0358160.
XX
PA   (PROT-) PROTEIN ENG CORP.
XX
PI   Guterman SK,  Kent RB,  Ladner RC,  Ley AC,  Markland W;
PI   Roberts BL;
XX
DR   WPI; 1996-321851/32.
DR   P-PSDB; AAR99213.
XX
PT   New engineered inhibitors of human neutrophil elastase - contg.
PT   aprotinin-like Kunitz domain for treating, e.g. cystic fibrosis or
PT   other respiratory disorders
XX
PS   Example 1; Page 43; 105pp; English.
XX
CC   Genetically engineered human derived Kunitz domains can be used to
CC   inhibit human neutrophil elastase, an enzyme involved in the
CC   elimination of pathogens and the restructuring of connective tissue.
CC   In cases of reduction of the circulating alpha-1-protease inhibitor
CC   (API or alphal antitrypsin), or the inactivation of API by oxidation
CC   (smokers emphysema), extensive destruction of the lung tissue may
CC   result from uncontrolled elastolytic activity of human neutrophil
CC   elastase. Other respiratory disorders such as cystic fibrosis are
CC   thought to be caused by human neutrophil elastase release by
CC   neutrophils.  The genetically engineered human derived Kunitz
CC   domains can be used to treat such respiratory disorders. See
CC   AAR99146-R99211. Fusion genes were used in the production of the
CC   Kunitz domain derivatives.  This sequence is a fragment of a fusion
CC   gene that encodes the M13 III signal peptide, ITI-D1 Kunitz domain
CC   and the mature M13 III protein. Substitutions in the ITI-D1 domain
CC   can be made by use of BglI, EagI, NcoI, StyI, PstI and KasI
CC   restriction sites found within the sequence or by
CC   oligonucleotide-directed mutagenesis of single-stranded phage DNA.
CC   The resulting peptide derivatives can be expressed as intact human
CC   neutrophil elastase binding proteins or can be incorporated into
CC   larger proteins as a domain.
XX
SQ   Sequence 285 BP; 74 A; 67 C; 68 G; 76 T; 0 other;


alignment_scores:
           Quality:  130.00              Length:     72
             Ratio:  3.333                 Gaps:      2
 Percent Similarity:  54.167   Percent Identity: 41.667

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAT35164   ..

 Align seg 1/1  to: AAT35164  from: 1  to: 285

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||      :::|||   |||:::|||      :::   |||   |||
      49 GGCGCCAAAGAAGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCAT 98

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
            :::      ||||||||||||||||   ||||||      |||||||
      99 GGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGA 148

      44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
         |||||   ::::::||||||   ||||||   |||||||||      :::
     149 CTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAA 198

      61 IleGlu..............ArgGluVal..................... 65
            |||               |||   |||
     199 AAGGAGTGTCTGCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGA 248

      66 .AlaCysValAlaLys 70
          :::|||:::||||||
     249 AAGTTGTTTAGCAAAA 264
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV40046

seq_documentation_block:
ID   AAV40046 standard; cDNA to mRNA; 441 BP.
XX
AC   AAV40046;
XX
DT   09-NOV-1998  (first entry)
XX
DE   rUTI cDNA SEQ ID NO:28 from WO9829453 Example 10.
XX
KW   Drug; cell membrane-directed drug; phospholipid; lipid bilayer;
KW   cell cortex; blood coagulation; inflammation; immunological disorder;
KW   ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..441
FT                   /*tag=  a
FT                   /note= "no stop codon given"
XX
PN   WO9829453-A1.
XX
PD   09-JUL-1998.
XX
PF   05-JAN-1998;   98WO-JP00002.
XX
PR   27-DEC-1996;   96JP-0359053.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Hasegawa T,  Kuriyama S;
XX
DR   WPI; 1998-388051/33.
DR   P-PSDB; AAW69522.
XX
PT   Drugs containing peptide(s) with specific affinity to
PT   phospholipid(s) - such as phosphatidyl serine, for treatment of
PT   blood coagulation, inflammatory and immunological disorders
XX
PS   Example 10; Page 84; 117pp; Japanese.
XX
CC   The present invention describes drug compositions which contain as
CC   an active component a peptide which has specific affinity to particular
CC   phospholipids (such as phosphatidyl ethanolamine or phosphatidyl
CC   serine), especially to phospholipids which constitute a lipid bilayer of
CC   cellular cortex and of which the concentration in the bilayer increases
CC   in cells which are abnormal (e.g. through injury, denaturation or
CC   activation). In particular, the peptide contains a sequence having
CC   phospholipid affinity and a structure of formula (I): (A1)a-(A2)b-(A3)c,
CC   where (A1) is one of two specific sequences (see AAW69516 and AAW69519),
CC   (A2) and (A3) are TRYLRIHPQSWVHQIALR, LRYLRIHPQSWVHQIALR (see AAW69517)
CC   or MEVLGCEAQNLY (see AAW69518); a = 0-5; b = 1-5, and c = 0-5. Preferred
CC   are the formulae A1-A2-A3, A2-A3, A2-A2-A3, A2-A2-A2-A3 or A2-A2
CC   (especially A2-A2-A3, A2-A2-A2-A3 or A2-A2). The sequence is linked to a
CC   peptide such as a blood factor, especially thrombo-modulin, urina-statin
CC   or membrane cofactor protein. The drugs are used for the treatment and
CC   prevention of diseases involving blood coagulation, inflammatory and
CC   immunological disorders. The present sequence encodes a protein from the
CC   present invention.
XX
SQ   Sequence 441 BP; 99 A; 119 C; 133 G; 90 T; 0 other;


alignment_scores:
            Quality:  128.50              Length:     66
              Ratio:  3.212                Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAV40046   ..
```

```
Align seg 1/1  to: AAV40046  from: 1  to: 441

      2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
        |||:::::::::::::::|||                 :::|||   |||::
     46 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 80

     18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
        :|||      :::   |||   |||      :::       ||||||||||
     81 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 130

     35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
        |||||   ||||||      ||||||||||||   ::::::||||||
    131 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 180

     52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
        ||||||   |||||||||      :::   |||         :::|||
    181 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 228

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1988.DAT:AAN81432

seq_documentation_block:
ID   AAN81432 standard; cDNA; 1232 BP.
XX
AC   AAN81432;
XX
DT   06-DEC-1990  (first entry)
XX
DE   Sequence encoding new fusion protein contg. alpha-1-microglobulin (AMG)
DE   and the HI-30 region of inter-alpha-trypsin inhibitor (III) light chain.
XX
KW   Serine protease; enzyme; pancreatitis; atherosclerosis;
KW   chronic inflammation; therapy; elastase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..1109
FT                   /*tag=  a
FT   mat_peptide     108..656
FT                   /*tag=  b
FT                   /product=AMG
FT   mat_peptide     666..1100
FT                   /*tag=  c
FT                   /product=HI-30
FT   misc_feature    921..926
FT                   /*tag=  d
FT                   /note="Differs from the protein sequence of HI-30
FT                         purified from urine"
FT   misc_feature    1077..1079
FT                   /*tag=  e
FT                   /note="Differs from the protein sequence of HI-30
FT                         purified from urine"
XX
PN   EP255011-A.
XX
PD   03-FEB-1988.
XX
PF   20-JUL-1987;   87EP-0110461.
XX
PR   29-JUL-1986;   86US-0891469.
XX
PA   (MILE ) MILES LABORATORIES INC.
XX
PI   Kaumeyer JF,  Kotick MP,  Polazzi JO;
XX
DR   WPI; 1988-030262/05.
DR   P-PSDB; AAP81110.
XX
PT   New DNA sequence coding for fusion protein contg. alpha-microglobulin -
PT   and inter-alpha-trypsin inhibitor, useful for treating excessive
PT   elastase prodn.
XX
```

```
PS   Disclosure; ; p; English.
XX
CC   A cDNA sequence encoding AMG in sequence with the HI-30 coding region of
CC   the light chain is claimed. The cDNA also contains a 5'-sequence coding
CC   for a secretory leader. 2 Arg residues separate the HI-30 and AMG regions
CC   and preceding AMG is a signal sequence. ITI is a serine protease,
CC   potentially used for treating excessive release of hydrolytic enzymes, eg
CC   elastase, in conditions such as pancreatitis, atherscleresis and chronic
CC   inflammation.
XX
SQ   Sequence 1232 BP; 299 A; 325 C; 364 G; 244 T; 0 other;


alignment_scores:
           Quality:  128.50            Length:     66
             Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAN81432   ..

 Align seg 1/1  to: AAN81432  from: 1  to: 1232

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||               :::|||   |||::
     711 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 745

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     746 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 795

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     796 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 845

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     846 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 893

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT78950

seq_documentation_block:
ID   AAT78950 standard; DNA; 1255 BP.
XX
AC   AAT78950;
XX
DT   11-NOV-1997  (first entry)
XX
DE   Anti-trypsin inhibitor UTI coding sequence.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             681..1124
FT                   /*tag= a
FT                   /product= protease inhibitor UTI
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25928.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 4-8; 37pp; Japanese.
XX
CC   This is the nucleotide sequence encoding the anti-trypsin inhibitor UTI.
CC   which is a protein containing 2 Kunitz domains.  The nucleotide sequence
CC   was used to construct a novel protease inhibitor by replacing the active
CC   site of the Kunitz domain 1 (see AAT79080, AAT79081 and AAT79083).
CC   The modified protease inhibitors are targeted to the protease elastase,
CC   especially from neutrophils and can be used to treat diseases associated
CC   with elastase.  Modifications of the active site were done by site
CC   directed mutagenesis.
XX
SQ   Sequence 1255 BP; 310 A; 329 C; 370 G; 246 T; 0 other;


alignment_scores:
           Quality:  128.50            Length:     66
             Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAT78950   ..

 Align seg 1/1  to: AAT78950  from: 1  to: 1255

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||               :::|||   |||::
     726 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 760

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     761 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 810

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     811 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 860

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     861 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 908

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77939

seq_documentation_block:
ID   AAC77939 standard; cDNA; 1444 BP.
XX
AC   AAC77939;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:333.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
```

```
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43730.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 885; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1444 BP; 362 A; 383 C; 422 G; 277 T; 0 other;


alignment_scores:
            Quality:  128.50              Length:     66
              Ratio:  3.212                Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAC77939   ..

 Align seg 1/1  to: AAC77939  from: 1  to: 1444

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||               :::|||   |||::
     895 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 929

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     930 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 979

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     980 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 1029

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
    1030 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 1077

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ06878

 seq_documentation_block:
 ID   AAQ06878 standard; DNA; 191 BP.
 XX
 AC   AAQ06878;
 XX
 DT   05-MAR-1991  (first entry)

XX
DE   Encodes human bikunin domain 1-derived protease inhibitor.
XX
KW   bikunin; serine protease inhibitor; rheumatoid arthritis; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   mutation        57..59
FT                   /*tag= a
FT                   /note= "substitute with codon for Leu, Ile, Val,
FT                   Arg, Phe, Tyr, Thr or Lys"
FT
FT   mutation        63..65
FT                   /*tag= b
FT                   /note= "substitute with codon for Leu, Arg, Ile,
FT                   Val or Lys"
FT
XX
PN   EP401508-A.
XX
PD   12-DEC-1990.
XX
PF   01-MAY-1990;   90EP-0108284.
XX
PR   18-JAN-1990;   90DE-4001244.
PR   13-MAY-1989;   89DE-3915689.
XX
PA   (FARB ) BAYER AG.
XX
PI   Fritz H,  Gebhard W,  Das R;
XX
DR   WPI; 1990-369615/50.
DR   P-PSDB; AAR08291.
XX
PT   Protease-inhibitor variant derived from human bikunin - prodn.
PT   used as therapeutic for treatment of eg. rheumatoid arthritis
XX
PS   Disclosure; Fig 2; 29pp; German.
XX
CC   The 5' end overhangs the 3' end of the complementary strand by 5
CC   bases; the 5' end of the complementary strand overhangs the 3' end
CC   of the sense strand by AGCTG. The sequence encodes residues 19-77 of
CC   mature bikunin in which Met(36) and/or Met(38) are substituted by
CC   another naturally occurring amino acid. The sequence comprises domain
CC   1 of bikunin and is synthesised from two synthetic oligonucleotides.
CC   The substitutions are achieved by site-directed mutagenesis.
CC   Plasmids containing the sequences are used to transform bacteria
CC   and the protease inhibitory fragments are secreted as fusion
CC   proteins. The fragments can be used for the treatment of emphysema,
CC   septic shock, rheumatoid arthritis and coagulation disorders.
CC   See also AAQ06879.
XX
SQ   Sequence 191 BP; 49 A; 43 C; 46 G; 53 T; 0 other;


alignment_scores:
            Quality:  127.00              Length:     54
              Ratio:  3.848                Gaps:      0
 Percent Similarity:  61.111   Percent Identity: 46.296

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAQ06878   ..

 Align seg 1/1  to: AAQ06878  from: 1  to: 191

      14 LysAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHi 30
         :::|||   |||:::|||      :::   |||   |||      :::
      18 GAAGACTCTTGCCAGCTGGGCTACTCTGCTGGTCCGTGCATGGGTATGAC 67

      30 sPheArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValP 47
            ||||||||||||||||   ||||||      ||||||||||||   :
      68 TTCTCGTTACTTCTACAACGGTACCTCTATGGCTTGCGAAACTTTCCAGT 117

      47 heSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArg 63
```

```
      :::::||||||   ||||||   |||||||||      :::   |||
    118 ACGGTGGTTGCATGGGTAACGGTAACAACTTCGTTACTGAAAAAGAATGC 167

     64 GluValAlaCys 67
           :::|||
    168 CTGCAGACTTGC 179

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ46936

seq_documentation_block:
ID   AAQ46936 standard; cDNA; 418 BP.
XX
AC   AAQ46936;
XX
DT   19-JAN-1994  (first entry)
XX
DE   cDNA encoding synthetic yeast leader and TFPI 2 variant.
XX
KW   TFPI; Kunitz sequence; pancreatitis; arthritis; proteolysis; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..409
FT                   /*tag=  a
FT                   /product= signal peptide and TFPI 2.
FT   sig_peptide     77..235
FT                   /*tag=  b
FT   mat_peptide     236..409
FT                   /*tag=  c
FT                   /product= TFPI 2.
XX
PN   WO9314121-A.
XX
PD   22-JUL-1993.
XX
PF   07-JAN-1993;   93WO-DK00004.
XX
PR   07-JAN-1992;   92WO-DK00001.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Bjorn SE,  Norris F,  Norris K,  Olsen O,  Petersen LC;
XX
DR   WPI; 1993-243149/30.
DR   P-PSDB; AAR39807.
XX
PT   New variants of human Kunitz protease inhibitor domain 2 - of
PT   tissue factor pathway inhibitor, with selective activity, for
PT   treating proteolytic diseases, e.g. pancreatitis or inflammation
XX
PS   Example 5; Page 42-43; 51pp; English.
XX
CC   Modification of the Kunitz sequence provides preferential inhibition
CC   of many proteases and so can be used to treat pathological
CC   proteolysis caused by enzyme release from polymorphonuclear
CC   leucocytes e.g. acute pancreatitis, inflammation, thrombocytopaenia,
CC   rheumatoid arthritis and psoriasis.  The variant can also be used
CC   for the preservation of organ and platelet function.
XX
SQ   Sequence 418 BP; 140 A; 71 C; 92 G; 115 T; 0 other;


alignment_scores:
           Quality:  127.00            Length:     63
             Ratio:  3.175               Gaps:      0
 Percent Similarity:  63.492   Percent Identity: 41.270

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAQ46936   ..

 Align seg 1/1  to: AAQ46936  from: 1  to: 418
```

```
      5 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 21
        :::||||||::::::               |||   |||   |||:::
    212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

     21 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 38
         :::   |||   |||            :::|||||||||||||||:::
    262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

     38 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 54
        :::|||:::|||||||   |||   ::::::||||||:::|||||||:::
    312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCAGGGGCAATATG 361

     55 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
        ||||||||||||         |||   :::      |||
    362 AACAATTTTAAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79081

seq_documentation_block:
ID   AAT79081 standard; DNA; 372 BP.
XX
AC   AAT79081;
XX
DT   11-NOV-1997  (first entry)
XX
DE   Ep7-d21-RPDF-52-55 coding sequence.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Synthetic.
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25933.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Page -; 37pp; Japanese.
XX
CC   This is the nucleotide sequence encoding a construct comprising the
CC   anti-trypsin inhibitor UTI Kunitz domains 1 and 2 (see AAT78950) linked
CC   at the 5' terminus to a sequence encoding the peptide Arg-Pro-Asp-Phe.
CC   The construct lacks the sequence encoding amino acids 1-21 and the
CC   Arg-Pro-Asp-Phe sequence replaces the native amino acids 22-26.  The
CC   construct also has the sequence encoding the active site site and the
CC   adjacent amino acids (amino acids 36-40 of the wild type sequence) and
CC   amino acids 52 and 55 substituted.  The modified protease inhibitors are
CC   targeted to the protease elastase, especially from neutrophils and can
CC   be used to treat diseases associated with elastase.  Modifications of
CC   the active site were done by site directed mutagenesis.
CC   Note: this sequence is not given in the specification but is constructed
CC   from the related construct Ep1-d21-RPDF-52-55 (AAT79080).
XX
SQ   Sequence 372 BP; 74 A; 98 C; 111 G; 89 T; 0 other;


alignment_scores:
           Quality:  126.00            Length:     53
             Ratio:  3.818               Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 47.170

alignment_block:
```

```
US-09-215-435-177_COPY_25_99 x AAT79081   ..

Align seg 1/1  to: AAT79081  from: 1  to: 372

     15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
        |||   |||:::|||      :::   |||   |||      :::
      7 GATTTCTGTCAACTGGGCTACTCGGCCGGTCCCTGCGTTGCTATGTTTCC 56

     31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||   ||||||      |||:::||||||||:::
     57 TAGGTATTTCTATAATGGTACATCCATGGCCTGTCAGACTTTCGTGTACG 106

     48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
        ::|||||   ||||||   |||||||||      :::   |||
    107 GCGGATGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTG 156

     65 ValAlaCys 67
           :::|||
    157 CAGACCTGC 165

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT79088

seq_documentation_block:
ID   AAT79088 standard; DNA; 467 BP.
XX
AC   AAT79088;
XX
DT   12-NOV-1997  (first entry)
XX
DE   Insert from plasmid pHH341.
XX
KW   Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW   neutrophil; disease; modification; site directed mutagenesis; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             7..441
FT                   /*tag=  a
FT                   /product= elastase inhibitor Ep7-d21-RPDF-52-55
FT   sig_peptide     7..66
FT                   /*tag=  b
FT   mat_peptide     67..438
FT                   /*tag=  c
XX
PN   JP09124700-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0288527.
XX
PR   07-NOV-1995;   95JP-0288527.
XX
PA   (GREC ) GREEN CROSS CORP.
XX
DR   WPI; 1997-316576/29.
DR   P-PSDB; AAW25938.
XX
PT   New protease inhibitor - useful for treating diseases involving
PT   elastase
XX
PS   Disclosure; Fig 38; 37pp; Japanese.
XX
CC   This is the nucleotide sequence of the insert for the plasmid pHH341
CC   which contains the sequence encoding the novel elastase specific
CC   inhibitor Ep7-d21-RPDF-52-55 (AAT79081). The inhibitor coding sequence
CC   is linked downstream of the yeast invertase (SUC2) signal peptide
CC   sequence. The modified protease inhibitors are targeted to the protease
CC   elastase, especially from neutrophils and can be used to treat diseases
CC   associated with elastase.  Modifications of the active site were done by
CC   site directed mutagenesis.
XX
SQ   Sequence 467 BP; 90 A; 121 C; 132 G; 124 T; 0 other;

alignment_scores:
           Quality:  126.00              Length:      53
             Ratio:  3.818                Gaps:       0
 Percent Similarity:  62.264   Percent Identity: 47.170

alignment_block:
 US-09-215-435-177_COPY_25_99 x AAT79088   ..

 Align seg 1/1  to: AAT79088  from: 1  to: 467

     15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
        |||   |||:::|||      :::   |||   |||      :::
     73 GATTTCTGTCAACTGGGCTACTCGGCCGGTCCCTGCGTTGCTATGTTTCC 122

     31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||   ||||||      |||:::||||||||:::
    123 TAGGTATTTCTATAATGGTACATCCATGGCCTGTCAGACTTTCGTGTACG 172

     48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
        ::|||||   ||||||   |||||||||      :::   |||
    173 GCGGATGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTG 222

     65 ValAlaCys 67
           :::|||
    223 CAGACCTGC 231
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-177_COPY_25_99 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  3:58 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-6 +  130.00  305.89  2.2e-09   1455  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-17 +  127.00  310.97  1.1e-09   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-15 +  123.00  300.66  4.3e-09   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-1 +  121.00  286.76  2.5e-08   979  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-1 +  121.00  286.76  2.5e-08   979  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-1 +  121.00  286.76  2.5e-08   979  !  S
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-817-145-2 +  121.00  285.20  3.1e-08   1140  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-13 +  120.00  292.92  1.1e-08   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-11 +  119.00  290.35  1.6e-08   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-135 +  117.00  290.51  1.6e-08   249  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-2 +  116.00  281.32  5.1e-08   474  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-4 +  116.00  277.42  8.4e-08   693  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-026-145-1 +  116.00  274.43  1.2e-07   928  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-446-646-8 +  116.00  274.43  1.2e-07   928  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-7 +  116.00  274.24  1.3e-07   945  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-854-764-1 +  116.00  273.01  1.5e-07   1065  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09377-1 +  116.00  273.01  1.5e-07   1065
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-054-782-1 +  116.00  269.98  2.2e-07   1431  !
/cgn1_7/ptodata/1/ina/backfiles1.seq:5466783-1 +  116.00  269.98  2.2e-07   1431  ! Pate
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-6 +  115.00  276.09  1.0e-07   614  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-14 +  114.00  287.00  2.5e-08   165  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-14 +  114.00  287.00  2.5e-08   165  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-14 +  114.00  287.00  2.5e-08   165  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-70 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-74 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-70 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-74 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-70 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-74 +  113.00  275.03  1.1e-07   412  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-084-718-58 +  111.00  269.73  2.3e-07   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-976-58 +  111.00  269.73  2.3e-07   418  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-443-977-58 +  111.00  269.73  2.3e-07   418  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-829-876-90 +  110.00  266.51  3.4e-07   445  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-985-692-1 +  107.50  261.18  6.7e-07   399  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-155-331-1 +  107.50  261.18  6.7e-07   399  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-424-022-1 +  107.50  261.18  6.7e-07   399  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-424-017B-1 +  107.50  261.18  6.7e-07   399  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-11696-1 +  107.50  261.18  6.7e-07   399
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-155-331-12 +  107.50  238.25  1.3e-05   3725  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-424-022-12 +  107.50  238.25  1.3e-05   3725  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-424-017B-12 +  107.50  238.25  1.3e-05   3725
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-11696-12 +  107.50  238.25  1.3e-05   3
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-829-876-92 +  107.00  258.78  9.2e-07   445
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-334-773A-5 +  106.00  256.84  1.2e-06   418
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-384-489-8 +  106.00  256.69  1.2e-06   424  !

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-6

seq_documentation_block:
; Sequence 6, Application US/08358160
; Patent No. 5663143
;  GENERAL INFORMATION:
;    APPLICANT:  LEY, Arthur C.
;    APPLICANT:  LADNER, Robert C.
;    APPLICANT:  GUTERMAN, Sonia K.
;    APPLICANT:  ROBERTS, Bruce L.
;    APPLICANT:  MARKLAND, William
;    APPLICANT:  KENT, Rachel B.
;    TITLE OF INVENTION:  ENGINEERED HUMAN-DERIVED KUNITZ
;    TITLE OF INVENTION:  DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  234
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W. Suite 300
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/358,160
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/133,031
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/009,319
;      FILING DATE:  26-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/664,989
;      FILING DATE:  01-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,063
;      FILING DATE:  02-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/240,160
;      FILING DATE:  02-SEP-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cooper, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  LEY=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1455 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  synthetic DNA fragment
US-08-358-160-6


alignment_scores:
            Quality:  130.00              Length:     72
```

```
            Ratio:  3.333                Gaps:      2
Percent Similarity:  54.167   Percent Identity: 41.667

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-358-160-6   ..

 Align seg 1/1  to: US-08-358-160-6  from: 1  to: 1455

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||     :::|||   |||:::|||      :::   |||   |||
      49 GGCGCCAAAGAAGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCAT 98

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
            :::      ||||||||||||||||   ||||||      |||||||
      99 GGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGA 148

      44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
         |||||   ::::::|||||||   ||||||   |||||||||      :::
     149 CTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAA 198

      61 IleGlu.............ArgGluVal..................... 65
            |||              |||   |||
     199 AAGGAGTGTCTGCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGA 248

      66 .AlaCysValAlaLys 70
          :::|||:::|||||
     249 AAGTTGTTTAGCAAAA 264

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-17

seq_documentation_block:
; Sequence 17, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-17


alignment_scores:
            Quality:  127.00              Length:      63
              Ratio:  3.175                Gaps:       0
Percent Similarity:  63.492   Percent Identity: 41.270

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-321-658B-17   ..

 Align seg 1/1  to: US-08-321-658B-17  from: 1  to: 418

       5 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 21
         :::||||||:::::                  |||   |||   |||:::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      21 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 38
          :::   |||   |||                :::||||||||||||||:::
     262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

      38 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 54
         :::|||:::|||||   |||   :::::::|||||||:::|||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCAGGGGCAATATG 361

      55 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||||||||         |||   :::       |||
     362 AACAATTTTAAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-15

seq_documentation_block:
; Sequence 15, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-15


alignment_scores:
            Quality:  123.00            Length:     63
              Ratio:  3.075              Gaps:      0
 Percent Similarity:  63.492   Percent Identity: 39.683

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-321-658B-15   ..

 Align seg 1/1  to: US-08-321-658B-15  from: 1  to: 418

       5 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 21
         :::||||||::::::               |||   |||   |||:::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      21 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 38
          :::   |||   |||          :::||||||||||||||||:::
     262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

      38 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 54
         :::|||:::||||||   |||   ::::::||||||:::||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCAGGGGCAATATG 361

      55 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         |||||||||:::         |||   :::      |||
     362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-147-710-1

seq_documentation_block:
; Sequence 1, Application US/08147710
; Patent No. 5455338
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  4225 Roosevelt Way, N.E.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,710
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-684
;      REFERENCE/DOCKET NUMBER:  93-14
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-547-8080 ext 322
;      TELEFAX:  206-548-2329
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Placenta
;    IMMEDIATE SOURCE:
;      CLONE:  J-2-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  39..746
US-08-147-710-1


alignment_scores:
            Quality:  121.00            Length:     70
              Ratio:  2.814              Gaps:      0
 Percent Similarity:  61.429   Percent Identity: 32.857

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-147-710-1   ..

 Align seg 1/1  to: US-08-147-710-1  from: 1  to: 979

       6 AlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAs 22
         |||::::::      |||:::      :::   |||   |||   :::::
     111 GCTCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGA 160

      22 nPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrS 39
         ::::|||   |||      :::   :::||||||:::|||:::|||   :
     161 CTACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACA 210

      39 erLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 55
         :::::   |||:::   |||:::::::::||||||:::||||||   |||
     211 CGCAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAAC 260

      56 AsnPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLy 72
         ||||||              :::   ||||||              ||
     261 AATTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAA 310

      72 sProProArg 75
         |   |||:::
     311 AGTTCCCAAA 320

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-458-090-1

seq_documentation_block:
; Sequence 1, Application US/08458090
; Patent No. 5728674
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,090
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  93-14D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Placenta
;    IMMEDIATE SOURCE:
;      CLONE:  J-2-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  39..746
US-08-458-090-1


alignment_scores:
            Quality:  121.00              Length:     70
              Ratio:  2.814                Gaps:      0
 Percent Similarity:  61.429   Percent Identity: 32.857

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-458-090-1    ..

 Align seg 1/1  to: US-08-458-090-1  from: 1  to: 979

       6 AlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAs 22
         |||::::::      |||:::      :::   |||   |||   :::::
     111 GCTCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGA 160

      22 nPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrS 39
         :::::|||   |||      :::   :::|||||||:::|||:::|||    :
     161 CTACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACA 210

      39 erLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 55
         :::::   |||:::   |||::::::::::|||||||:::|||||   |||
     211 CGCAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAAC 260

      56 AsnPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLy 72
         ||||||                :::   ||||||             ||
     261 AATTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAA 310

      72 sProProArg 75
         |   |||:::
     311 AGTTCCCAAA 320

 seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-457-887-1

 seq_documentation_block:

; Sequence 1, Application US/08457887
; Patent No. 5914315
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Kisiel, Walter
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  NOVEL HUMAN KUNITZ-TYPE INHIBITORS AND
;    TITLE OF INVENTION:  METHODS RELATING THERETO
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,887
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  93-14D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Placenta
;    IMMEDIATE SOURCE:
;      CLONE:  J-2-11
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  39..746
US-08-457-887-1


alignment_scores:
            Quality:  121.00              Length:     70
              Ratio:  2.814                Gaps:      0
 Percent Similarity:  61.429   Percent Identity: 32.857

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-457-887-1    ..

 Align seg 1/1  to: US-08-457-887-1  from: 1  to: 979

       6 AlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAs 22
         |||::::::      |||:::      :::   |||   |||   :::::
     111 GCTCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGA 160

      22 nPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrS 39
         :::::|||   |||      :::   :::|||||||:::|||:::|||    :
     161 CTACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACA 210

      39 erLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 55
         :::::   |||:::   |||::::::::::|||||||:::|||||   |||
     211 CGCAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAAC 260

      56 AsnPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLy 72
```

```
          ||||||                  :::   ||||||           ||
     261 AATTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAA 310

      72 sProProArg 75
         |   |||:::
     311 AGTTCCCAAA 320

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-817-145-2

seq_documentation_block:
; Sequence 2, Application US/08817145
; Patent No. 6025329
;  GENERAL INFORMATION:
;    APPLICANT:  UTSUMI, Jun
;    APPLICANT:  SUDO, Tetsuo
;    APPLICANT:  TANAKA, Yasuhiko
;    APPLICANT:  MATSUI, Mizuo
;    TITLE OF INVENTION:  THERAPEUTIC AGENT FOR OPHTHALMIC
;    TITLE OF INVENTION:  DISEASES
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch, LLP.
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/817,145
;      FILING DATE:  02-JUL-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  760-230P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..753
US-08-817-145-2


alignment_scores:
            Quality:  121.00            Length:     70
              Ratio:  2.814               Gaps:      0
 Percent Similarity:  61.429   Percent Identity: 32.857

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-817-145-2   ..

 Align seg 1/1  to: US-08-817-145-2  from: 1  to: 1140

       6 AlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAs 22
         |||::::::      |||:::     :::   |||   |||   :::::
     121 GCTCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAGA 170

      22 nPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrS 39
         :::::|||   |||      :::   :::||||||:::|||:::|||   :
     171 CTACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACA 220

      39 erLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 55
         :::::   |||:::   |||:::::::::||||||:::||||||   |||
     221 CGCAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAAC 270

      56 AsnPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLy 72
         ||||||           :::   ||||||           ||
     271 AATTTCTACACCTGGGAGGCTTGCGACGATGCTTGCTGGAGGATAGAAAA 320

      72 sProProArg 75
         |   |||:::
     321 AGTTCCCAAA 330

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-13

seq_documentation_block:
; Sequence 13, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-13


alignment_scores:
            Quality:  120.00            Length:     63
```

```
            Ratio:  3.077                Gaps:      0
Percent Similarity:  61.905   Percent Identity: 39.683

alignment_block:
US-09-215-435-177_COPY_25_99 x US-08-321-658B-13   ..

Align seg 1/1  to: US-08-321-658B-13  from: 1  to: 418

       5 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 21
         :::||||||::::::                |||   |||   |||::::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      21 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 38
          :::   |||   |||              :::||||||||||||||:::
     262 AGATCCTGGAATATGTAAAGCTCGTATTATCAGGTATTTTTATAACAATC 311

      38 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 54
         :::|||:::||||||   |||   ::::::||||||   ||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCCTGGGCAATATG 361

      55 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||||||:::         |||   :::     |||
     362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-11

seq_documentation_block:
; Sequence 11, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  418 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic/human
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..409
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  77..235
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  236..409
US-08-321-658B-11


alignment_scores:
            Quality:  119.00              Length:     63
              Ratio:  3.132                 Gaps:      0
Percent Similarity:  60.317   Percent Identity: 39.683

alignment_block:
US-09-215-435-177_COPY_25_99 x US-08-321-658B-11   ..

Align seg 1/1  to: US-08-321-658B-11  from: 1  to: 418

       5 IleAlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMe 21
         :::||||||::::::                |||   |||   |||::::
     212 ATGGCTGAGAGATTGGAGAAGAGAAAGCCAGATTTCTGCTTTTTGGAAGA 261

      21 tAsnPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgT 38
          :::   |||   |||                ||||||||||||||:::
     262 AGATCCTGGAATATGTCGAGGTTATATTACCAGGTATTTTTATAACAATC 311

      38 hrSerLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeu 54
         :::|||:::||||||   |||   ::::::||||||   ||||||:::
     312 AGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCCTGGGCAATATG 361

      55 AsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||||||:::         |||   :::     |||
     362 AACAATTTTGAGACACTGGAAGAATGCAAGAACATTTGT 400

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-358-160-135

seq_documentation_block:
; Sequence 135, Application US/08358160
; Patent No. 5663143
;  GENERAL INFORMATION:
;    APPLICANT:  LEY, Arthur C.
;    APPLICANT:  LADNER, Robert C.
;    APPLICANT:  GUTERMAN, Sonia K.
;    APPLICANT:  ROBERTS, Bruce L.
;    APPLICANT:  MARKLAND, William
;    APPLICANT:  KENT, Rachel B.
;    TITLE OF INVENTION:  ENGINEERED HUMAN-DERIVED KUNITZ
;    TITLE OF INVENTION:  DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  234
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W. Suite 300
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/358,160
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/133,031
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/009,319
;      FILING DATE:  26-JAN-1993
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 07/664,989
;      FILING DATE:  01-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,063
;      FILING DATE:  02-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/240,160
;      FILING DATE:  02-SEP-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cooper, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  LEY=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  135:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  249 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  synthetic DNA fragment
US-08-358-160-135


alignment_scores:
            Quality:  117.00              Length:     57
              Ratio:  3.441                 Gaps:      0
 Percent Similarity:  59.649   Percent Identity: 42.105

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-358-160-135   ..

 Align seg 1/1  to: US-08-358-160-135  from: 1  to: 249

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||         |||   |||   |||:::   :::   |||   |||
      49 GGCGCCAAGCCTGACTTCTGCTTCCTCGAGGAGGATCCCGGGATTTGCCG 98

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
                 ||||||||||||||||:::   :::|||:::||||||
      99 CGGTTATATTACGCGTTATTTCTATAATAACCAGACTAAGCAATGTGAGC 148

      44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
          |||   ::::::|||||   ||||||:::|||||||||:::
     149 GGTTCAAGTATGGTGGTTGCCTAGGTAATATGAACAACTTCGAGACTCTA 198

      61 IleGluArgGluValAlaCys 67
            |||   :::      |||
     199 GAAGAGTGTAAGAACATATGT 219

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-2

seq_documentation_block:
; Sequence 2, Application US/07828920A
; Patent No. 5312736
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Jesper
;    APPLICANT:  No. 5312736dfang, Ole Juul
;    TITLE OF INVENTION:  Anticoagulant Protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5312736o No. 5312736disk of No. 5312736th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6200
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/828,920A
;      FILING DATE:  19920127
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 4080/89
;      FILING DATE:  18-AUG-1989
;      APPLICATION NUMBER:  WO PCT/DK90/00212
;      FILING DATE:  17)AUG)1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zelson, Steve T.
;      REGISTRATION NUMBER:  30335
;      REFERENCE/DOCKET NUMBER:  3287.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..437
US-07-828-920A-2


alignment_scores:
            Quality:  116.00              Length:     53
              Ratio:  3.515                 Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 43.396

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-07-828-920A-2   ..

 Align seg 1/1  to: US-07-828-920A-2  from: 1  to: 474

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||   |||:::   :::   |||   |||
     234 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 283

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
          ||||||||||||||:::   :::|||:::||||||   |||   ::::
     284 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 333

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||:::|||||||||:::          |||   :::
     334 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 383

      65 ValAlaCys 67
            |||
     384 AACATTTGT 392

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-828-920A-4

seq_documentation_block:
; Sequence 4, Application US/07828920A
; Patent No. 5312736
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Jesper
;    APPLICANT:  No. 5312736dfang, Ole Juul
;    TITLE OF INVENTION:  Anticoagulant Protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5312736o No. 5312736disk of No. 5312736th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6200
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
```

```
;       ZIP:  10174-6201
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/828,920A
;       FILING DATE:  19920127
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  DK 4080/89
;       FILING DATE:  18-AUG-1989
;       APPLICATION NUMBER:  WO PCT/DK90/00212
;       FILING DATE:  17)AUG)1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Zelson, Steve T.
;       REGISTRATION NUMBER:  30335
;       REFERENCE/DOCKET NUMBER:  3287.204-US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212 867 0123
;       TELEFAX:  212 867 0298
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  693 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  105..656
US-07-828-920A-4


alignment_scores:
            Quality:  116.00              Length:      53
              Ratio:  3.515                Gaps:       0
 Percent Similarity:  62.264   Percent Identity: 43.396

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-07-828-920A-4    ..

 Align seg 1/1  to: US-07-828-920A-4  from: 1  to: 693

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||   |||:::   :::   |||   |||
     453 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 502

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
          ||||||||||||||:::   :::|||:::||||||   |||   ::::
     503 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 552

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||:::|||||||||:::         |||   :::
     553 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 602

      65 ValAlaCys 67
             |||
     603 AACATTTGT 611

 seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-026-145-1

 seq_documentation_block:
 ; Sequence 1, Application US/08026145
 ; Patent No. 5378614
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Petersen, Jens G. Litske
 ;    APPLICANT:  No. 5378614dfang, Ole Juul
 ;    TITLE OF INVENTION:  Method for Making TFPI Analogues
 ;    NUMBER OF SEQUENCES:  6
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  No. 5378614o No. 5378614disk of No. 5378614th America, Inc.
 ;      STREET:  405 Lexington Avenue, Suite 6200
 ;      CITY:  New York
 ;      STATE:  N. Y.
 ;      COUNTRY:  United States of America
 ;      ZIP:  10174-6201
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/026,145
 ;      FILING DATE:  19930302
 ;      CLASSIFICATION:  435
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  US 07/828,920
 ;      FILING DATE:  27-JAN-1992
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  WO PCT/DK/90/00212
 ;      FILING DATE:  17-AUG-1990
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  DK 4080/89
 ;      FILING DATE:  18-AUG-1989
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Agris, Cheryl H.
 ;      REGISTRATION NUMBER:  34086
 ;      REFERENCE/DOCKET NUMBER:  3321.214-US
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  212 867 0123
 ;      TELEFAX:  212 867 0298
 ;  INFORMATION FOR SEQ ID NO:  1:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  928 base pairs
 ;      TYPE:  NUCLEIC ACID
 ;      STRANDEDNESS:  double
 ;      TOPOLOGY:  linear
 ;    MOLECULE TYPE:  DNA
 ;    HYPOTHETICAL:  NO
 ;    ANTI-SENSE:  NO
 ;    ORIGINAL SOURCE:
 ;      ORGANISM:  Synthetic
 ;    FEATURE:
 ;      NAME/KEY:  CDS
 ;      LOCATION:  8..919
 ;    FEATURE:
 ;      NAME/KEY:  sig_peptide
 ;      LOCATION:  8..91
 ;    FEATURE:
 ;      NAME/KEY:  mat_peptide
 ;      LOCATION:  92..919
 US-08-026-145-1


 alignment_scores:
             Quality:  116.00              Length:      53
               Ratio:  3.515                Gaps:       0
  Percent Similarity:  62.264   Percent Identity: 43.396

 alignment_block:
  US-09-215-435-177_COPY_25_99 x US-08-026-145-1    ..

  Align seg 1/1  to: US-08-026-145-1  from: 1  to: 928

       15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
          |||   |||   |||:::   :::   |||   |||
      374 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 423

       31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
           ||||||||||||||:::   :::|||:::||||||   |||   ::::
      424 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 473

       48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
          ::||||||   ||||||:::|||||||||:::         |||   :::
```

```
     474 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 523

      65 ValAlaCys 67
                 |||
     524 AACATTTGT 532

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-446-646-8

seq_documentation_block:
; Sequence 8, Application US/08446646
; Patent No. 5726038
;  GENERAL INFORMATION:
;    APPLICANT:  Christiansen, Lars
;    APPLICANT:  Petersen, Jens G.
;    TITLE OF INVENTION:  A DNA Construct Encoding the YAP3 Signal
;    TITLE OF INVENTION:  Peptide
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5726038o No. 5726038disk of No. 5726038th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,646
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3987.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..919
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  8..91
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  92..919
US-08-446-646-8


alignment_scores:
            Quality:  116.00              Length:     53
              Ratio:  3.515                 Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 43.396

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-446-646-8   ..

 Align seg 1/1  to: US-08-446-646-8  from: 1  to: 928
```

```
      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||   |||:::   :::   |||   |||
     374 GATTTCTGCTTTTTGGAAGAGGATCCTGGAATATGTCGAGGTTATATTAC 423

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||:::   :::|||:::||||||   |||   ::::
     424 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAAAGGTTCAAGTATG 473

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||:::||||||||||:::         |||   :::
     474 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTCGAGGAATGCAAG 523

      65 ValAlaCys 67
                 |||
     524 AACATTTGT 532

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-321-658B-7

seq_documentation_block:
; Sequence 7, Application US/08321658B
; Patent No. 5576294
;  GENERAL INFORMATION:
;    APPLICANT:  No. 5576294ris, Fanny
;    APPLICANT:  No. 5576294ris, Kjeld
;    APPLICANT:  Bjorn, Soren E
;    APPLICANT:  Petersen, Lars C
;    APPLICANT:  Olsen, Ole H
;    TITLE OF INVENTION:  A HUMAN KUNITZ-TYPE PROTEASE INHIBITOR
;    TITLE OF INVENTION:  VARIANT
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5576294o No. 5576294disk of No. 5576294th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,658B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney Dr., Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  3693.210-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  945 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  365..538
US-08-321-658B-7


alignment_scores:
            Quality:  116.00              Length:     53
              Ratio:  3.515                 Gaps:      0
 Percent Similarity:  62.264   Percent Identity: 43.396
```

```
alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-321-658B-7   ..

 Align seg 1/1  to: US-08-321-658B-7  from: 1  to: 945

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||   |||:::   :::   |||   |||
     371 GATTTCTGCTTTTTGGAAGAAGATCCTGGAATATGTCGAGGTTATATTAC 420

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
          ||||||||||||||:::   :::|||:::||||||   |||   ::::
     421 CAGGTATTTTTATAACAATCAGACAAAACAGTGTGAACGTTTCAAGTATG 470

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::||||||   ||||||:::|||||||||:::          |||   :::
     471 GTGGATGCCTGGGCAATATGAACAATTTTGAGACACTGGAAGAATGCAAG 520

      65 ValAlaCys 67
             |||
     521 AACATTTGT 529
```

```
OM of: US-09-215-435-177_COPY_25_99 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799 +  419.00 1013.36  4.2e-47   499
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799 +  419.00 1013.36  4.2e-47   49
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76 +  419.00 1012.84  4.5e-47   526
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303 +  409.00  988.06  1.1e-45   539
/cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136 +  409.00  988.06  1.1e-45   539
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-516-448-1671 +  368.00  891.74  2.5e-40   34
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-833-381-1671 +  368.00  891.74  2.5e-40   349
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-843-620-654 +  368.00  891.74  2.5e-40   349
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-188-162-1870 -  335.00  805.64  1.6e-35   58
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148 +  307.00  744.54  4.0e-32   271
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-208-69337 +  251.00  606.72  1.9e-24   2
/cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-131-380-189 +  210.00  509.11  5.2e-19   208
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-499-23002 +  210.00  509.11  5.2e-19   2
/cgn1_7/ptodata/1/pna/US6005_COMB.seq:US-60-055-705-189 +  210.00  509.11  5.2e-19   208
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-6802 +  134.00  316.03  2.9e-08   404
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-960-352-6802 +  134.00  316.03  2.9e-08   404
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-13925 +  134.00  315.67  3.1e-08   41
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-960-352-13925 +  134.00  315.67  3.1e-08   41
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-14462A-23 +  133.00  267.51  1.5e-05   425
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-01313-23 +  133.00  267.51  1.5e-05   4252
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-235-810-23 +  133.00  267.51  1.5e-05   42521
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658-23 +  133.00  267.51  1.5e-05   42521
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658B-23 +  133.00  267.51  1.5e-05   4252
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-373-658C-23 +  133.00  267.51  1.5e-05   4252
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-171-503-23 +  133.00  267.51  1.5e-05   4252
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6889 +  131.00  307.64  8.6e-08   44
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-6788 +  131.00  305.87  1.1e-07   53
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-24855 +  131.00  291.05  7.2e-07   2
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-142-896-264 -  131.00  290.52  7.7e-07   253
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-191-637-6787 +  131.00  290.49  7.7e-07   25
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6888 -  131.00  289.20  9.1e-07   28
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-167-217-6885 -  131.00  284.45  1.7e-06   46
/cgn1_7/ptodata/1/pna/US6015_COMB.seq:US-60-150-584-381 +  131.00  277.87  3.9e-06   908
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-237A-859 -  131.00  247.60  0.0002   19
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-849-406-3 +  130.00  309.66  6.6e-08   285  1
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-203-3184 +  129.50  302.23  1.7e-07   53
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-190-812 +  129.50  302.21  1.7e-07   535
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-160-203-2095 +  129.50  302.21  1.7e-07   53
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-192-739-1393 +  129.50  299.02  2.6e-07   73
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-143-992-158 -  129.00  285.37  1.5e-06   259
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-331-793-28 +  128.50  301.66  1.8e-07   44
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-235-076-30403 +  128.50  300.19  2.2e-07
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-289-768-37096 +  128.50  300.19  2.2e-07
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-332-782-30403 +  128.50  300.19  2.2e-07
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-737-223-30403 +  128.50  300.19  2.2e-07

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5799

seq_documentation_block:
; Sequence 5799, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 5799
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 22..318
;   NAME/KEY: sig_peptide
;   LOCATION: 22..93
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
US-09-834-366-5799


alignment_scores:
            Quality:  419.00              Length:     75
              Ratio:  5.587                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-834-366-5799   ..

 Align seg 1/1  to: US-09-834-366-5799  from: 1  to: 499

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      94 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 143

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     144 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTWCACTTTAGATATT 193

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     194 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCGGCTGT 243

      51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     244 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAGCCTG 293

      67 sValAlaLysTyrLysProProArg 75
         |||||||||||||||||||||||||
     294 TGTTGCAAAATACAAACCACCGAGG 318

seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5799

seq_documentation_block:
; Sequence 5799, Application US/60197873
```

```
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
;  FILE REFERENCE:  81.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/197,873
;  CURRENT FILING DATE:  2000-04-18
;  NUMBER OF SEQ ID NOS:  52153
;  SOFTWARE:  Patent.pm
; SEQ ID NO 5799
;    LENGTH: 499
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 22..318
;    NAME/KEY: sig_peptide
;    LOCATION: 22..93
;    OTHER INFORMATION: Von Heijne matrix
;    OTHER INFORMATION: score 4.59999990463257
;    OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
US-60-197-873-5799


alignment_scores:
            Quality:  419.00              Length:      75
              Ratio:  5.587                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-60-197-873-5799   ..

 Align seg 1/1  to: US-60-197-873-5799  from: 1  to: 499

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      94 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 143

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     144 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTWCACTTTAGATATT 193

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     194 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCGGCTGT 243

      51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     244 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAGCCTG 293

      67 sValAlaLysTyrLysProProArg 75
         |||||||||||||||||||||||||
     294 TGTTGCAAAATACAAACCACCGAGG 318

 seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-76

 seq_documentation_block:
 ; Sequence 76, Application US/09215435A
 ; GENERAL INFORMATION:
 ;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
 ;  APPLICANT: Duclert Aymeric,
 ;  APPLICANT: Bougueleret Lydie
 ;  TITLE OF INVENTION: Extended cDNAs
 ;  FILE REFERENCE: GENSET.019A
 ;  CURRENT APPLICATION NUMBER: US/09/215,435A
 ;  CURRENT FILING DATE: 1998-12-17
 ;  EARLIER APPLICATION NUMBER: 60/069,957
 ;  EARLIER FILING DATE: 1997-12-17
 ;  EARLIER APPLICATION NUMBER: 60/074,121
 ;  EARLIER FILING DATE: 1998-2-9
 ;  EARLIER APPLICATION NUMBER: 60/081,563
 ;  EARLIER FILING DATE: 1998-4-13
 ;  EARLIER APPLICATION NUMBER: 60/096,116
 ;  EARLIER FILING DATE: 1998-8-10
 ;  NUMBER OF SEQ ID NOS: 519
 ;  SOFTWARE: Patent.pm
 ; SEQ ID NO 76
 ;    LENGTH: 526
 ;    TYPE: DNA
 ;    ORGANISM: Homo sapiens
 ;    FEATURE:
 ;    NAME/KEY: CDS
 ;    LOCATION: 22..318
 ;    FEATURE:
 ;    NAME/KEY: sig_peptide
 ;    LOCATION: 22..93
 ;    OTHER INFORMATION: Von Heijne matrix
 ;    OTHER INFORMATION: score 4.6
 ;    OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
 ;    FEATURE:
 ;    NAME/KEY: polyA_signal
 ;    LOCATION: 497..502
 ;    FEATURE:
 ;    NAME/KEY: polyA_site
 ;    LOCATION: 516..526
 US-09-215-435-76


 alignment_scores:
             Quality:  419.00              Length:      75
               Ratio:  5.587                Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-177_COPY_25_99 x US-09-215-435-76   ..

  Align seg 1/1  to: US-09-215-435-76  from: 1  to: 526

        1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       94 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 143

       17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      144 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 193

       34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      194 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCGGCTGT 243

       51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      244 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAGCCTG 293

       67 sValAlaLysTyrLysProProArg 75
          |||||||||||||||||||||||||
      294 TGTTGCAAAATACAAACCACCGAGG 318

  seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-303

  seq_documentation_block:
  ; Sequence 303, Application US/09215435A
  ; GENERAL INFORMATION:
  ;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
  ;  APPLICANT: Duclert Aymeric,
  ;  APPLICANT: Bougueleret Lydie
  ;  TITLE OF INVENTION: Extended cDNAs
  ;  FILE REFERENCE: GENSET.019A
  ;  CURRENT APPLICATION NUMBER: US/09/215,435A
  ;  CURRENT FILING DATE: 1998-12-17
  ;  EARLIER APPLICATION NUMBER: 60/069,957
  ;  EARLIER FILING DATE: 1997-12-17
  ;  EARLIER APPLICATION NUMBER: 60/074,121
```

```
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 303
;   LENGTH: 539
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 32..328
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 32..103
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.59999990463257
;   OTHER INFORMATION: seq FFIFCSLNTLLLG/GV
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 508..513
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 528..539
US-09-215-435-303


alignment_scores:
            Quality:  409.00              Length:     75
              Ratio:  5.453                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.333

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-215-435-303   ..

 Align seg 1/1  to: US-09-215-435-303  from: 1  to: 539

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     104 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 153

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 203

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||:::|||
     204 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 253

      51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     254 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAKCCTG 303

      67 sValAlaLysTyrLysProProArg 75
         |||||||||||||||||||||||||
     304 TGTTGCAAAATACAAACCACCGAGG 328

seq_name: /cgn1_7/ptodata/1/pna/US6006_COMB.seq:US-60-069-957-136

seq_documentation_block:
; Sequence 136, Application US/60069957
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  EXTENDED cDNAS FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  381
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/069,957
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.019PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  136:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  32..328
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  508..513
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  528..539
;      IDENTIFICATION METHOD:  blastn
US-60-069-957-136


alignment_scores:
            Quality:  409.00              Length:     75
              Ratio:  5.453                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.333

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-60-069-957-136   ..

 Align seg 1/1  to: US-60-069-957-136  from: 1  to: 539

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     104 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 153

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 203

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||:::|||
     204 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 253

      51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     254 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAKCCTG 303
```

```
     67 sValAlaLysTyrLysProProArg 75
        |||||||||||||||||||||||||
    304 TGTTGCAAAATACAAACCACCGAGG 328

seq_name: /cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-516-448-1671

seq_documentation_block:
; Sequence 1671, Application US/09516448
; GENERAL INFORMATION:
;  APPLICANT:  Robison, Keith E.
;  TITLE OF INVENTION:  Novel Nucleic Acid and Protein Homologs
;  FILE REFERENCE:  5800-119
;  CURRENT APPLICATION NUMBER: US/09/516,448
;  CURRENT FILING DATE:  2000-02-29
;  NUMBER OF SEQ ID NOS:  2050
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1671
;   LENGTH: 349
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(349)
;   OTHER INFORMATION: n = A,T,C or G
US-09-516-448-1671


alignment_scores:
            Quality:  368.00              Length:     68
              Ratio:  5.662                 Gaps:      0
 Percent Similarity:  95.588   Percent Identity: 95.588

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-516-448-1671   ..

 Align seg 1/1  to: US-09-516-448-1671  from: 1  to: 349

       8 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 24
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

      24 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 41
         |||||||||||||||||||||||||||||||||||||||||||||||   |
      73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122

      41 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 57
         |||||||||||||||   |||||||||||||||||||||||||||||||
     123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

      58 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 74
         ||||||||||||   |||||||||||||||||||||||||||||||||||
     173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

      74 oArg 75
         ||||
     223 GAGG 226

 seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-833-381-1671

 seq_documentation_block:
 ; Sequence 1671, Application US/09833381
 ; GENERAL INFORMATION:
 ;  APPLICANT: Robison, Keith E.
 ;  TITLE OF INVENTION: Novel Nucleic Acid and Protein Homologs
 ;  FILE REFERENCE: 5800-119
 ;  CURRENT APPLICATION NUMBER: US/09/833,381
 ;  CURRENT FILING DATE:  2001-04-11
 ;  PRIOR APPLICATION NUMBER: 09/516,448
 ;  PRIOR FILING DATE: 2000-02-29
 ;  NUMBER OF SEQ ID NOS: 2050
 ;  SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 1671
 ;   LENGTH: 349
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: misc_feature
 ;   LOCATION: (1)...(349)
 ;   OTHER INFORMATION: n = A,T,C or G
 US-09-833-381-1671


 alignment_scores:
             Quality:  368.00              Length:     68
               Ratio:  5.662                 Gaps:      0
  Percent Similarity:  95.588   Percent Identity: 95.588

 alignment_block:
  US-09-215-435-177_COPY_25_99 x US-09-833-381-1671   ..

  Align seg 1/1  to: US-09-833-381-1671  from: 1  to: 349

        8 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 24
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

       24 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 41
          |||||||||||||||||||||||||||||||||||||||||||||||   |
       73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122

       41 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 57
          |||||||||||||||   |||||||||||||||||||||||||||||||
      123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

       58 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 74
          ||||||||||||   |||||||||||||||||||||||||||||||||||
      173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

       74 oArg 75
          ||||
      223 GAGG 226

 seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-843-620-654

 seq_documentation_block:
 ; Sequence 654, Application US/09843620
 ; GENERAL INFORMATION:
 ;  APPLICANT: Gearing, David P.
 ;  APPLICANT:  Holtzman, Douglas A.
 ;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ;  TITLE OF INVENTION:  THEREFOR
 ;  FILE REFERENCE: 1600.2098-001
 ;  CURRENT APPLICATION NUMBER: US/09/843,620
 ;  CURRENT FILING DATE:  2001-04-26
 ;  PRIOR APPLICATION NUMBER: US 60/200,099
 ;  PRIOR FILING DATE: 2000-04-27
 ;  NUMBER OF SEQ ID NOS: 1135
 ;  SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 654
 ;   LENGTH: 349
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: misc_feature
 ;   LOCATION: (1)...(349)
 ;   OTHER INFORMATION: n = A,T,C or G
 US-09-843-620-654


 alignment_scores:
             Quality:  368.00              Length:     68
               Ratio:  5.662                 Gaps:      0
  Percent Similarity:  95.588   Percent Identity: 95.588

 alignment_block:
  US-09-215-435-177_COPY_25_99 x US-09-843-620-654   ..
```

```
 Align seg 1/1  to: US-09-843-620-654  from: 1  to: 349

       8 LysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGl 24
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      23 AAGATATGTGGAGACCTCAAAGATCCCTGCAAATTGGACATGAATTTTGG 72

      24 ySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysA 41
         |||||||||||||||||||||||||||||||||||||||||||||   |
      73 AAGCTGCTATGAAGTTCACTTTAGATATTTCTACAACAGAACCTCCANAA 122

      41 rgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPhe 57
         |||||||||||||   |||||||||||||||||||||||||||||||||
     123 GATGTGAAACTTTTGNCTTCTCCGGCTGTAATGGCAACCTTAACAACTTC 172

      58 LysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLysProPr 74
         ||||||||||||   |||||||||||||||||||||||||||||||||||
     173 AAGCTTAAAATANAACGTGAAGTAGCCTGTGTTGCAAAATACAAACCACC 222

      74 oArg 75
         ||||
     223 GAGG 226

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-188-162-1870

seq_documentation_block:
; Sequence 1870, Application US/60188162
; GENERAL INFORMATION:
;  APPLICANT:  Beasley, Ellen
;  TITLE OF INVENTION:  ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL000518
;  CURRENT APPLICATION NUMBER: US/60/188,162
;  CURRENT FILING DATE:  2000-05-09
;  NUMBER OF SEQ ID NOS:  5094
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1870
;   LENGTH: 585
;   TYPE: DNA
;   ORGANISM: HUMAN
US-60-188-162-1870


alignment_scores:
            Quality:  335.00             Length:     60
              Ratio:  5.583               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.333

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-60-188-162-1870/rev   ..

 Align seg 1/1  to reverse of: US-60-188-162-1870  from: 1  to: 585

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 GATCCCTGCAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTT 207

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 TAGATATTTCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCT 157

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ||:::||||||||||||||||||||||||||||||||||||||||||||||
     156 CCAGCTGTAATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAA 107

      65 ValAlaCysValAlaLysTyrLysProPro 74
         ||||||||||||||||||||||||||||||
     106 GTAGCCTGTGTTGCAAAATACAAACCACCG 77

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-148

seq_documentation_block:
; Sequence 148, Application US/08905279
;  GENERAL INFORMATION:
;    APPLICANT:  Genset SA
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,279
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.011A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  148:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  32..103
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 4.6
;      OTHER INFORMATION:  seq FFIFCSLNTLLLG/GV
US-08-905-279-148


alignment_scores:
            Quality:  307.00             Length:     55
              Ratio:  5.582               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.182

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-905-279-148   ..

 Align seg 1/1  to: US-08-905-279-148  from: 1  to: 271

       1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     104 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 153

      17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 203

      34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
         ||||||||||||||||||||||||||||||||||||||||||:::|||||
     204 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCAGCTGT 253

      51 AsnGlyAsnLeuAsn 55
         |||||||||||||||
     254 AATGGCAACCTTAAC 268
```

```
seq_name: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-208-69337

seq_documentation_block:
; Sequence 69337, Application US/09540208
; GENERAL INFORMATION:
;  APPLICANT:  Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES OF MALE REPRODUCTIVE TISSUE
;  FILE REFERENCE:  PD-1029 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,208
;  CURRENT FILING DATE:  2000-03-31
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS:  70811
;  SOFTWARE:  PERL Program
; SEQ ID NO 69337
;   LENGTH: 282
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: rat00103589
;   NAME/KEY: unsure
;   LOCATION: 22, 255
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-208-69337


alignment_scores:
            Quality:  251.00              Length:     71
              Ratio:  4.328                Gaps:      0
 Percent Similarity:  81.690   Percent Identity: 59.155

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-540-208-69337   ..

 Align seg 1/1  to: US-09-540-208-69337  from: 1  to: 282

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::|||:::|||   :::   |||   :::::::::|||||||||
       1 GTGGAGAAGGTTGCTGCAAGGNTCTGCAAAGAATTCAACGATCCTTGTGC 50

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
          |||:::|||:::   ||||||||||||||||   |||:::|||:::||||
      51 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 100

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||:::|||||||||:::|||:::   ||||||||||||||||||||
     101 ACAACCAAACCTCCAAAAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAAT 150

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysVa 68
         ||||||   |||||||||   ||||||||||:::   :::|||   |||
     151 GGTAACCGTAATAACTTCATGCTTAAAATAGACTGTGACGTAATGTGTAA 200

      68 lAlaLysTyrLys 72
         :::||||||
     201 TGAAGAATACAAG 213

 seq_name: /cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-131-380-189

 seq_documentation_block:
 ; Sequence 189, Application US/09131380
 ; GENERAL INFORMATION:
 ;  APPLICANT: Gooding, Douglas H.
 ;  APPLICANT: Stuve, Laura L.
 ;  APPLICANT: Stuart, Susan G.
 ;  APPLICANT: Naughton, Rebecca E.
 ;  APPLICANT: Garrow, Bonnie L.
 ;  APPLICANT: Klemm, Juli D.
 ;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM RAT FEMUR
 ;  FILE REFERENCE: PZ-0013 US
 ;  CURRENT APPLICATION NUMBER: US/09/131,380
 ;  CURRENT FILING DATE: 1998-08-07
 ;  EARLIER APPLICATION NUMBER: 60/055,705
 ;  EARLIER FILING DATE: 1997-08-08
 ;  EARLIER APPLICATION NUMBER: 60/074,278
 ;  EARLIER FILING DATE: 1998-02-10
 ;  NUMBER OF SEQ ID NOS: 2199
 ;  SOFTWARE: PERL Program
 ; SEQ ID NO 189
 ;   LENGTH: 208
 ;   TYPE: DNA
 ;   ORGANISM: Rattus norvegicus
 ;   FEATURE:
 ;   NAME/KEY: unsure
 ;   LOCATION: 109, 117, 156, 167
 ;   OTHER INFORMATION: a or g or c or t, unknown, or other
 ;   FEATURE: -
 ;   OTHER INFORMATION: 700292496H1
 US-09-131-380-189


 alignment_scores:
             Quality:  210.00              Length:     67
               Ratio:  4.038                Gaps:      1
  Percent Similarity:  77.612   Percent Identity: 55.224

 alignment_block:
  US-09-215-435-177_COPY_25_99 x US-09-131-380-189   ..

  Align seg 1/1  to: US-09-131-380-189  from: 1  to: 208

        2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
          |||:::|||:::|||   :::::::|||   :::::::::|||||||||
        2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

       18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
           |||:::|||:::   ||||||||||||||||   |||:::|||:::||||
       52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

       35 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 51
          |||||   ||||||    :::|||:::   ||||||||||||||||||
      102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

       51 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
          ||||      |||   |||   ||||||||||:::   :::|||   |||
      152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200

  seq_name: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-499-23002

  seq_documentation_block:
  ; Sequence 23002, Application US/09540499
  ; GENERAL INFORMATION:
  ;  APPLICANT:  Seilhamer, Jeffrey J.
  ;  APPLICANT:  Delegeane, Angelo M.
  ;  APPLICANT:  Stuart, Susan G.
  ;  APPLICANT:  Stuve, Laura L.
  ;  APPLICANT:  Mullahy, Sara J.
  ;  APPLICANT:  Naughton, Rebecca E.
  ;  TITLE OF INVENTION:  POLYNUCLEOTIDES OF MUSCULOSKELETAL SYSTEM TISSUE
  ;  FILE REFERENCE:  PD-1032 CIP
  ;  CURRENT APPLICATION NUMBER: US/09/540,499
  ;  CURRENT FILING DATE:  2000-03-31
  ;  Prior application data removed - refer to PALM or file wrapper
  ;  NUMBER OF SEQ ID NOS:  29520
  ;  SOFTWARE:  PERL Program
  ; SEQ ID NO 23002
  ;   LENGTH: 208
  ;   TYPE: DNA
  ;   ORGANISM: Rattus norvegicus
  ;   FEATURE:
  ;   NAME/KEY: misc_feature
```

```
;   OTHER INFORMATION: Incyte ID No: rat00198915
;   NAME/KEY: unsure
;   LOCATION: 109, 117, 156, 167
;   OTHER INFORMATION: a, t, c, g, or other
US-09-540-499-23002


alignment_scores:
            Quality:  210.00              Length:     67
              Ratio:  4.038                Gaps:      1
 Percent Similarity:  77.612   Percent Identity: 55.224

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-540-499-23002   ..

 Align seg 1/1  to: US-09-540-499-23002  from: 1  to: 208

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::|||:::|||   ::::::|||   ::::::::::||||||||
       2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
          |||:::|||:::   |||||||||||||||   |||:::|||:::||||
      52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

      35 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 51
         |||||   ||||||    :::|||:::   ||||||||||||||||||
     102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

      51 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||      |||   |||   ||||||||::   :::|||   |||
     152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200

seq_name: /cgn1_7/ptodata/1/pna/US6005_COMB.seq:US-60-055-705-189

seq_documentation_block:
; Sequence 189, Application US/60055705
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Argentine, Charles C.
;    APPLICANT:  Naughton, Rebecca E.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM RAT
;    TITLE OF INVENTION:  FEMUR
;    NUMBER OF SEQUENCES:  1093
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/055,705
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PZ-0013 P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 845-4166
;  INFORMATION FOR SEQ ID NO:  189:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  700292496H1
US-60-055-705-189


alignment_scores:
            Quality:  210.00              Length:     67
              Ratio:  4.038                Gaps:      1
 Percent Similarity:  77.612   Percent Identity: 55.224

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-60-055-705-189   ..

 Align seg 1/1  to: US-60-055-705-189  from: 1  to: 208

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::|||:::|||   ::::::|||   ::::::::::||||||||
       2 GTGGAGAAGGTTGCTGCAAGGCTCTGCAAAGAATTCAACGATCCTTGTGC 51

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
          |||:::|||:::   |||||||||||||||   |||:::|||:::||||
      52 TTTGGAAATGGATCCTGGCAGCTGCTATGAAAGCCATTACAGATTTTTCT 101

      35 yrAsnArgThrSer.LysArgCysGluThrPheValPheSerGlyCysAs 51
         |||||   ||||||    :::|||:::   ||||||||||||||||||
     102 ACAACCANACCTCCANCCAAGTGTCAGATTTTCGTGTTCAGTGGTTGCAA 151

      51 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||      |||   |||   ||||||||::   :::|||   |||
     152 TGGTNACCGTAATAANTTCATGCTTAAAATAGACTGTGACGTAATGTGT 200

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-480-902-6802

seq_documentation_block:
; Sequence 6802, Application US/09480902
; GENERAL INFORMATION:
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Warren, Wesley C.
;  TITLE OF INVENTION: Nucleic Acid and Other Molecules Associated with Lactation and
;  TITLE OF INVENTION: Muscle and Fat Deposition
;  FILE REFERENCE: 10298/1
;  CURRENT APPLICATION NUMBER: US/09/480,902
;  CURRENT FILING DATE: 2000-01-11
;  EARLIER APPLICATION NUMBER: USSN 60/115,707
;  EARLIER FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6802
;   LENGTH: 404
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 29-LIB34-029-Q1-E1-H1
US-09-480-902-6802


alignment_scores:
            Quality:  134.00              Length:     43
              Ratio:  4.467                Gaps:      0
 Percent Similarity:  69.767   Percent Identity: 55.814

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-480-902-6802   ..

 Align seg 1/1  to: US-09-480-902-6802  from: 1  to: 404

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         |||   |||:::||||||   :::   |||   |||       :::
     239 GATTCCTGCCAGCTGGACTACTCACAAGGGCCTTGCCTGGGGCTGTTCAA 288
```

```
 31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
    ||||||||||||||||  ||||||     ||||||||||||:::::::
289 GAGGTATTTCTATAACGGTACATCCATGGCCTGTGAGACCTTCTTATATG 338

 48 erGlyCysAsnGlyAsnLeuAsnAsnPhe 57
    ::||||||  ||||||  |||||||||
339 GCGGCTGCATGGGGAATGGCAACAACTTC 367
```

```
OM of: US-09-215-435-177_COPY_25_99 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:08 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-3 +  130.00  317.61  1.8e-10   28
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-78 +  117.00  286.01  1.0e-08    2
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-84 +  105.00  257.73  3.8e-07    2
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-80 +  100.00  245.68  1.8e-06    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11942 +   99.50  234.05  8.0e-06
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-1 +   98.50  238.13  4.8e-06   27
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-149 +   92.50  191.21  0.0020     6
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-9347 +   82.50  193.29  0.0015
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-82 +   81.00  197.55  0.0009     1
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-73 +   78.00  181.48  0.0068     4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11256 +   76.00  177.43  0.0114
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8534 +   75.00  174.48  0.0167
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-71 +   74.00  142.00  1.07       8
/cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-75 +   74.00  142.00  1.08       8
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-830-446-15 +   68.00  139.85  1.42      23
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8985 -   67.50  155.34  0.1943
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8984 -   67.50  154.45  0.2179
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-871-161-390 +   67.50  151.97  0.2995     5
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8986 -   62.50  142.04  1.07
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-768-185A-1 +   61.50   74.31  6.3e+03   32
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-871-161-93 -   60.50  133.53  3.19      64
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-7012 +   60.00  134.85  2.69
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-87 +   60.00  102.31  174.84    13
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2359 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2360 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2361 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2362 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2363 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2364 +   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2365 -   60.00  102.31  174.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-178 +   59.50  107.83  86.14      6
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-381 +   59.00  133.93  3.03       4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-303-24 -   58.50  113.68  40.65     28
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6503 +   58.00  117.12  26.16
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-7119 -   58.00  113.73  40.37
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3481 -   57.50  127.78  6.66
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3482 -   57.50  125.24  9.23
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-967-805-8 -   57.50  118.49  21.94     135
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-189 -   57.50  116.59  27.98      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8956 +   57.00  129.91  5.07
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-23571 +   57.00  129.80  5.14
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-9314 +   57.00  129.65  5.24
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-9548 +   57.00  128.95  5.73
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-834-975-955 +   56.50  112.22  49.05
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6903 +   56.50  111.30  55.15

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-3

seq_documentation_block:
; Sequence 3, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 285 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-849-406A-3


alignment_scores:
            Quality:  130.00              Length:     72
              Ratio:  3.333                Gaps:      2
 Percent Similarity:  54.167    Percent Identity: 41.667

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-3   ..

 Align seg 1/1  to: US-08-849-406A-3  from: 1  to: 285

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||      :::|||   |||:::|||      :::   |||   |||
      49 GGCGCCAAAGAAGACTCTTGCCAGCTGGGCTACTCGGCCGGTCCCTGCAT 98

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
            :::      ||||||||||||||||   ||||||      |||||||
      99 GGGAATGACCAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGA 148
```

```
     44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
        |||||   ::::::|||||||   |||||||   |||||||||      :::
    149 CTTTCCAGTACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAA 198

     61 IleGlu..............ArgGluVal..................... 65
          |||                 |||   |||
    199 AAGGAGTGTCTGCAGACCTGCCGAACTGTGGGCGCCGCTGAAACTGTTGA 248

     66 .AlaCysValAlaLys 70
         :::|||:::||||||
    249 AAGTTGTTTAGCAAAA 264

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-78

seq_documentation_block:
; Sequence 78, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 78:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 249 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-08-849-406A-78


alignment_scores:
            Quality:  117.00            Length:     57
              Ratio:  3.441               Gaps:      0
 Percent Similarity:  59.649   Percent Identity: 42.105

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-78   ..

 Align seg 1/1  to: US-08-849-406A-78  from: 1  to: 249
```

```
     11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
        |||          |||   |||   |||:::   :::   |||   |||
     49 GGCGCCAAGCCTGACTTCTGCTTCCTCGAGGAGGATCCCGGGATTTGCCG 98

     27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
                    ||||||||||||||||:::   :::|||:::||||||
     99 CGGTTATATTACGCGTTATTTCTATAATAACCAGACTAAGCAATGTGAGC 148

     44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
        |||   ::::::|||||||   |||||||:::||||||||:::
    149 GGTTCAAGTATGGTGGTTGCCTAGGTAATATGAACAACTTCGAGACTCTA 198

     61 IleGluArgGluValAlaCys 67
          |||   :::      |||
    199 GAAGAGTGTAAGAACATATGT 219

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-84

seq_documentation_block:
; Sequence 84, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 84:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 201 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-08-849-406A-84


alignment_scores:
            Quality:  105.00            Length:     57
              Ratio:  3.387               Gaps:      0
 Percent Similarity:  54.386   Percent Identity: 36.842

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-84   ..
```

```
 Align seg 1/1  to: US-08-849-406A-84  from: 1  to: 201

      11 GlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCysTy 27
         |||          |||   |||   |||:::         |||:::|||
       1 GGCGCCAAGCCTGACTTCTGCTTCCTCGAGGAGNRTVVSGGGMNTTGCRT 50

      27 rGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysGluT 44
                    |||   |||||||||:::   :::|||:::|||   :
      51 TGSTNWTTTTMNSCGTTDSTTCTATAATAACCAGGCTAAGCAATGTSWGV 100

      44 hrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLys 60
         ::|||   ::::::||||||      |||   |||||||||:::
     101 NATTCVHATATGGTGGTTGCVHGGSTAATVBGAACAACTTCGAGACTCTA 150

      61 IleGluArgGluValAlaCys 67
            |||   :::      |||
     151 GAAGAGTGTAAGAACATATGT 171

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-80

seq_documentation_block:
; Sequence 80, Application US/08849406A
;       APPLICANT: LEY, Arthur C.
;                  LADNER, Robert C.
;                  GUTERMAN, Sonia K.
;                  ROBERTS, Bruce L.
;                  MARKLAND, William
;                  KENT, Rachel B.
;       TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                           DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;       NUMBER OF SEQUENCES: 86
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;            STREET: 624 Ninth Street, N.W. Suite 300
;            CITY: Washington
;            STATE: District of Columbia
;            COUNTRY: USA
;            ZIP: 20001
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/849,406A
;            FILING DATE: 21-Jul-1999
;            CLASSIFICATION: 530
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/US95/16349
;            FILING DATE: 15-DEC-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Cooper, Iver P.
;            REGISTRATION NUMBER: 28,005
;            REFERENCE/DOCKET NUMBER: LEY=1A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 202-628-5197
;            TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 80:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 189 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-08-849-406A-80


alignment_scores:
           Quality:  100.00            Length:     51
             Ratio:  3.125               Gaps:      0
 Percent Similarity:  62.745   Percent Identity: 33.333

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-80   ..

 Align seg 1/1  to: US-08-849-406A-80  from: 1  to: 189

      17 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 33
         |||   :::      :::   |||   |||      :::      |||::
      22 TGCGCTTTCAAAGCTGATGACGGTCCGTGTAAAGCTATCATGAAACGTTT 71

      33 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 50
         :|||:::|||      ::::::::::||||||   |||::::::::::|||||
      72 CTTCTTCAACATTTTCACGCGTCAGTGCGAGGAATTCATTTACGGTGGTT 121

      50 ysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAla 66
         ||:::||||||   |||:::|||:::       |||   :::
     122 GTGAAGGTAACCAGAACCGGTTCGAATCTCTAGAGGAATGTAAGAAGATG 171

      67 Cys 67
         |||
     172 TGC 174

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11942

seq_documentation_block:
; Sequence 11942, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11942
;   LENGTH: 538
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(538)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-11942


alignment_scores:
           Quality:   99.50            Length:     59
             Ratio:  3.109               Gaps:      2
 Percent Similarity:  54.237   Percent Identity: 38.983

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-922-340-11942   ..

 Align seg 1/1  to: US-09-922-340-11942  from: 1  to: 538

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   ||||||   |||                        |||   ||
     331 TGCGTGGACCTGCCAGACACA.....................GGACTCTG 359

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |   |||        |||::::::||||||      |||::::::|||
     360 CAAGGAGAGCATCCCGCGCTGGTACTACAACCCCTTCAGTGAACACTGCG 409

      43 luThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeu 59
            |||::::::::::||||||   ||||||   |||||||||
     410 CCCGCTTTACCTATGGTGGTTGTTATGGCAACAAGAACAACTTT..... 453

      60 LysIleGluArgGluValAlaCysVal 68
```

```
              |||:::|||      |||:::
    454 ......GAGGAAGAGCAGCAGTGCCTC 474

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-1

seq_documentation_block:
; Sequence 1, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 276 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-849-406A-1


alignment_scores:
            Quality:   98.50              Length:     51
              Ratio:  3.078                Gaps:      2
 Percent Similarity:  62.745   Percent Identity: 43.137

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-1   ..

 Align seg 1/1  to: US-08-849-406A-1  from: 1  to: 276

      24 GlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLy 40
         |||   |||           :::|||||||||||||||      :::
      88 GGGCCCTGCAAAGCGCGCATCATCCGCTATTTCTACAATGCTAAAGCAGG 137

      40 sArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsnP 57
            |||:::|||||||||:::::|||||||:::::::::   |||||||
     138 CCTGTGCCAGACCTTTGTATACGGTGGTTGCCGTGCTAAGCGTAACAACT 187

      57 heLysLeuLysIleGluArgGlu.ValAlaCysValAlaLysTyrLysPr 73
         |||||         :::   ||| :::|||||||||            ||

     188 TTAAA.........TCGGCCGAAGATTGCATGCGTACCTGCGGTGGCGCC 228

      73 o 73
         |
     229 G 229

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-149

seq_documentation_block:
; Sequence 149, Application US/09971392
; GENERAL INFORMATION:
;  APPLICANT: Peterson, David P.
;  APPLICANT:  Pearson, Cecelia I.
;  APPLICANT:  Cocks, Benjamin G.
;  TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
;  FILE REFERENCE: PA-0029 US
;  CURRENT APPLICATION NUMBER: US/09/971,392
;  CURRENT FILING DATE:  2001-10-03
;  PRIOR APPLICATION NUMBER: 60/237,652
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 260
;  SOFTWARE: PERL Program
; SEQ ID NO 149
;   LENGTH: 6777
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Template ID: 235164.25
US-09-971-392-149


alignment_scores:
            Quality:   92.50              Length:     68
              Ratio:  2.312                Gaps:      2
 Percent Similarity:  58.824   Percent Identity: 26.471

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-971-392-149   ..

 Align seg 1/1  to: US-09-971-392-149  from: 1  to: 6777

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::::::::::|||:::::::      :::
     982 GTGGAAGAGGTGGTTCGAGAGGTGTGCTCTGAACAAGCCGAGACG..... 1026

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
                        |||   |||      :::      |||:::::::
    1027 ...............GGGCCGTGCCGAGCAATGATCTCCCGCTGGTACT 1060

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         :::::   |||:::   :::|||      |||   ::::::||||||:::
    1061 TTGATGTGACTGAAGGGAAGTGTGCCCCATTCTTTTACGGCGGATGTGGC 1110

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysVa 68
         ||||||   |||||||||            :::   |||      |||::
    1111 GGCAACCGGAACAACTTT............GACACAGAAGAGTACTGCAT 1148

      68 lAla 69
         :|||
    1149 GGCC 1152

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-9347

 seq_documentation_block:
 ; Sequence 9347, Application US/09922340
 ; GENERAL INFORMATION:
 ;  APPLICANT: Hyseq, Inc.
 ;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
 ;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
 ;  FILE REFERENCE: 20411-726CON2
 ;  CURRENT APPLICATION NUMBER: US/09/922,340
 ;  CURRENT FILING DATE:  2001-08-03
```

```
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9347
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(426)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-9347


alignment_scores:
            Quality:   82.50              Length:     49
              Ratio:  3.056                Gaps:      2
 Percent Similarity:  55.102   Percent Identity: 40.816

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-922-340-9347   ..

 Align seg 1/1  to: US-09-922-340-9347  from: 1  to: 426

      10 CysGlyAspLeuLysAspProCysLysLeuAspMetAsnPheGlySerCy 26
         |||   ||||||   |||                         |||   ||
     234 TGCGTGGACCTGCCAGACACA.....................GGACTCTG 262

      26 sTyrGluValHisPheArgTyrPheTyrAsnArgThrSerLysArgCysG 43
         |   |||         |||::::::||||||      |||::::::|||
     263 CAAGGAGAGCATCCCGCGCTGGTACTACAACCCCTTCAGCGAACACTGCG 312

      43 luThrPheValPheSerGlyCysAsnGlyAsn.LeuAsnAsnPhe 57
            |||::::::::::||||||   ||||||    |||||||||
     313 CCCGCTTTACCTATGGTGGTTGTTACGGNAACAANGAACAACTTT 357

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-82

seq_documentation_block:
; Sequence 82, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY=1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 82:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 189 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-08-849-406A-82


alignment_scores:
            Quality:   81.00              Length:     51
              Ratio:  3.000                Gaps:      0
 Percent Similarity:  52.941   Percent Identity: 29.412

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-82   ..

 Align seg 1/1  to: US-08-849-406A-82  from: 1  to: 189

      17 CysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTy 33
         |||   :::           |||:::|||              |||
      22 TGCGCTTTCAAAGCTNRTRVSGGTCNTTGTRTTGSTNTCTTCMNSCGTTD 71

      33 rPheTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyC 50
          |||:::|||      :::::::::|||   :::|||   ::::::||||
      72 STTCTTCAACATTTTCACGCGTCAGTGCSWGVHATTCVHATACGGTGGTT 121

      50 ysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAla 66
         ||      |||   |||:::|||:::         |||   :::
     122 GTVHGGSTAACSRGAACCGGTTCGAATCTCTAGAGGAATGTAAGAAGATG 171

      67 Cys 67
         |||
     172 TGC 174

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-73

seq_documentation_block:
; Sequence 73, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
```

```
;            FILING DATE: 15-DEC-1995
;         ATTORNEY/AGENT INFORMATION:
;            NAME: Cooper, Iver P.
;            REGISTRATION NUMBER: 28,005
;            REFERENCE/DOCKET NUMBER: LEY=1A
;         TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 202-628-5197
;            TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 73:
;         SEQUENCE CHARACTERISTICS:
;            LENGTH: 444 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-08-849-406A-73


alignment_scores:
            Quality:   78.00            Length:     61
              Ratio:  2.294              Gaps:      0
 Percent Similarity:  55.738   Percent Identity: 27.869

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-73   ..

 Align seg 1/1  to: US-08-849-406A-73  from: 1  to: 444

       7 GluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPh 23
         |||   :::      |||   ::::::   |||:::|||   :::
     241 GAAGGTGTTTCCTTGGACAAGAGAGAGGCTTGTAACTTGCCAATCGTCAG 290

      23 eGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerL 40
          |||   |||              |||:::   ::::::   ::::::
     291 AGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGTTAAGG 340

      40 ysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsn 56
         :::|||      |||   ::::::|||||:::|||||   |||:::
     341 GTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTAACAAG 390

      57 PheLysLeuLysIleGluArgGluValAlaCys 67
         |||      :::   |||   :::      |||
     391 TTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 423

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11256

seq_documentation_block:
; Sequence 11256, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11256
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-340-11256


alignment_scores:
            Quality:   76.00            Length:     34
              Ratio:  3.455              Gaps:      1
 Percent Similarity:  64.706   Percent Identity: 44.118
```

```
alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-922-340-11256   ..

 Align seg 1/1  to: US-09-922-340-11256  from: 1  to: 401

      35 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs 51
         ||||||      |||:::::|||      |||:::::::::||||||
       2 TACAACCCCTTCAGTGAACACTGCGCCCGCTTTACCTATGGTGGTTGTTA 51

      51 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCysV 68
          ||||||   |||||||||            |||:::|||      |||:
      52 TGGCAACAAGAACAACTTT...........GAGGAAGAGCAGCAGTGCC 89

      68 al 68
         ::
      90 TC 91

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8534

seq_documentation_block:
; Sequence 8534, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8534
;   LENGTH: 418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(418)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-8534


alignment_scores:
            Quality:   75.00            Length:     33
              Ratio:  3.409              Gaps:      0
 Percent Similarity:  66.667   Percent Identity: 42.424

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-922-340-8534   ..

 Align seg 1/1  to: US-09-922-340-8534  from: 1  to: 418

      35 TyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAs 51
         ||||||      |||:::::|||      |||:::::::::||||||
       2 TACAACCCCTTCAGTGAACACTGCGCCCGCTTTACCTATGGGGGTTGTTA 51

      51 nGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
          ||||||   |||||||||:::   |||   :::          :::|||
      52 TGGCAACAAGAACAACTTTGAGGAAAAGCAGCAGTGCCTCGAGTCTTGT 100

 seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-71

 seq_documentation_block:
 ; Sequence 71, Application US/08849406A
 ;         APPLICANT: LEY, Arthur C.
 ;                    LADNER, Robert C.
 ;                    GUTERMAN, Sonia K.
 ;                    ROBERTS, Bruce L.
 ;                    MARKLAND, William
```

```
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY-1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8584 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-08-849-406A-71


alignment_scores:
            Quality:   74.00            Length:     61
              Ratio:  2.242              Gaps:      0
 Percent Similarity:  54.098   Percent Identity: 27.869

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-71   ..

 Align seg 1/1  to: US-08-849-406A-71  from: 1  to: 8584

       7 GluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPh 23
         |||   :::      |||   :::      |||:::|||   :::
    1185 GAAGGTGTTTCCTTGGACAAGAGAGCTGCTTGTAACTTGCCAATCGTCAG 1234

      23 eGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerL 40
          |||   |||               |||:::   ::::::   ::::::
    1235 AGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGTTAAGG 1284

      40 ysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsn 56
         :::|||      |||   ::::::||||||:::||||||   |||:::
    1285 GTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTAACAAG 1334

      57 PheLysLeuLysIleGluArgGluValAlaCys 67
         |||      :::   |||   :::      |||
    1335 TTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 1367

seq_name: /cgn1_7/ptodata/1/pna/US08_NEW_COMB.seq:US-08-849-406A-75

seq_documentation_block:
; Sequence 75, Application US/08849406A
;        APPLICANT: LEY, Arthur C.
;                   LADNER, Robert C.
;                   GUTERMAN, Sonia K.
;                   ROBERTS, Bruce L.
;                   MARKLAND, William
;                   KENT, Rachel B.
;        TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
;                            DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;        NUMBER OF SEQUENCES: 86
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 624 Ninth Street, N.W. Suite 300
;             CITY: Washington
;             STATE: District of Columbia
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/849,406A
;             FILING DATE: 21-Jul-1999
;             CLASSIFICATION: 530
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US95/16349
;             FILING DATE: 15-DEC-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Cooper, Iver P.
;             REGISTRATION NUMBER: 28,005
;             REFERENCE/DOCKET NUMBER: LEY-1A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;    INFORMATION FOR SEQ ID NO: 75:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8590 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-08-849-406A-75


alignment_scores:
            Quality:   74.00            Length:     61
              Ratio:  2.242              Gaps:      0
 Percent Similarity:  54.098   Percent Identity: 27.869

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-08-849-406A-75   ..

 Align seg 1/1  to: US-08-849-406A-75  from: 1  to: 8590

       7 GluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAsnPh 23
         |||   :::      |||   :::      |||:::|||   :::
    1185 GAAGGTGTTTCCTTGGACAAGAGAGCTGCTTGTAACTTGCCAATCGTCAG 1234

      23 eGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrSerL 40
          |||   |||               |||:::   ::::::   ::::::
    1235 AGGTCCATGCATTGCTTTCTTCCCAAGATGGGCTTTCGACGCTGTTAAGG 1284

      40 ysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsnAsn 56
         :::|||      |||   ::::::||||||:::||||||   |||:::
    1285 GTAAGTGCGTCTTGTTCCCATACGGTGGTTGTCAAGGTAACGGTAACAAG 1334

      57 PheLysLeuLysIleGluArgGluValAlaCys 67
         |||      :::   |||   :::      |||
    1335 TTCTACTCTGAGAAGGAGTGTAGAGAGTACTGT 1367

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-830-446-15

seq_documentation_block:
```

```
; Sequence 15, Application US/09830446
; GENERAL INFORMATION:
;  APPLICANT: MURDIN, Andrew D.; OOMEN, Raymond P; and WANG, Joe
;  TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
;  FILE REFERENCE: 77813-51
;  CURRENT APPLICATION NUMBER: US/09/830,446
;  CURRENT FILING DATE:  2001-09-21
;  PRIOR APPLICATION NUMBER: US 60/106034
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: US 60/106039
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: US 60/10604
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: US 60/106044
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: US 60/106072
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 60/106073
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 60/106074
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 60/106087
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 60/106587
;  PRIOR FILING DATE: 1998-11-02
;  PRIOR APPLICATION NUMBER: US 60/106588
;  PRIOR FILING DATE: 1998-11-02
;  PRIOR APPLICATION NUMBER: US 60/107089
;  PRIOR FILING DATE: 1998-11-02
;  PRIOR APPLICATION NUMBER: US 60/107034
;  PRIOR FILING DATE: 1998-11-02
;  PRIOR APPLICATION NUMBER: US 60/107035
;  PRIOR FILING DATE: 1998-11-02
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 2300
;   TYPE: DNA
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(2194)
US-09-830-446-15


alignment_scores:
            Quality:   68.00              Length:     58
              Ratio:  2.061                Gaps:      2
 Percent Similarity:  56.897   Percent Identity: 32.759

alignment_block:
 US-09-215-435-177_COPY_25_99 x US-09-830-446-15    ..

 Align seg 1/1  to: US-09-830-446-15  from: 1  to: 2300

      23 PheGly.........SerCysTyrGluValHisPheArgTyrPheTyr.. 35
         ||||||         ::|||   :::   ::::::||||||   :::
    1522 TTTGGGAGGCTTATTAAGTGCCTTAGATCAAGATGGAGATACTGTCATTG 1571

      36 .............AsnArgThrSerLysArgCysGluThrPheValPheS 48
                     |||:::   |||      :::   :::||||||||||
    1572 TCTTGAATCCTAGAATCATGGCTCAGGATACGCAACAAGCTTCGTTTTTT 1621

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
            :::|||||||   |||   |||:::::::::         |||:::
    1622 GTAGGGCAAACGGTCCCTTACCAAACTATCAAATACTATATCCAAGAAAC 1671

      65 ValAlaCysValAlaLysTyrLys 72
               |||   |||||||||:::
    1672 AGGAACTGTAACGCAAAATATCGA 1695
```

OM of: US-09-215-435-177_COPY_25_99 to: EST:* out_format : pfs

Date: Nov 14, 2001 10:30 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-177_COPY_25_99
Query length: 75
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est45:AW303354	-	419.00	1021.98	8.3e-48	443	! AW303354 xv18b01.x1 Soares_NFL_
gb_htc:AK020315	+	269.00	647.96	5.7e-27	611	! AK020315 Mus musculus adult mal
gb_gss32:AZ763929	-	235.00	564.55	2.5e-22	570	! AZ763929 1M0559C11R Mouse 10kb
gb_est15:AI015730	-	132.00	311.20	3.2e-08	496	! AI015730 ot71b08.s1 Soares_tota
gb_est18:AI256949	+	131.00	307.70	5.1e-08	552	! AI256949 LP04037.5prime LP Dros
gb_est104:D83828	+	129.00	302.75	9.6e-08	552	! D83828 HUMHBC4837 Human pancrea
gb_est109:T74655	+	128.50	305.91	6.4e-08	350	! T74655 yc57e11.r1 Stratagene li
gb_est7:AA424642	+	128.50	304.31	7.8e-08	413	! AA424642 zv91d02.r1 Soares_NhHM
gb_est14:AF063548	+	128.50	302.26	1.0e-07	511	! AF063548 AF063548 Homo sapiens
gb_est16:AI110674	+	128.50	302.26	1.0e-07	511	! AI110674 HA0104 Human fetal liv
gb_est40:AV720273	-	128.50	302.09	1.0e-07	520	! AV720273 AV720273 GLC Homo sapi
gb_est17:AI207711	+	128.50	302.07	1.0e-07	521	! AI207711 HA3093 Human fetal liv
gb_est17:AI207590	-	128.50	301.19	1.2e-07	571	! AI207590 HA2954 Human fetal liv
gb_est40:AV658823	+	128.50	300.30	1.3e-07	626	! AV658823 AV658823 GLC Homo sapi
gb_est25:AI807643	-	128.50	300.03	1.4e-07	644	! AI807643 wf49d04.x1 Soares_NFL_
gb_est15:AI065076	+	128.50	299.76	1.4e-07	662	! AI065076 HA0910 Human fetal liv
gb_est40:AV651758	+	128.50	299.63	1.4e-07	671	! AV651758 AV651758 GLC Homo sapi
gb_est43:AW173259	-	128.50	299.57	1.4e-07	675	! AW173259 xj85c08.x1 Soares_NFL_
gb_est40:AV700236	-	128.50	299.40	1.5e-07	687	! AV700236 AV700236 GKC Homo sapi
gb_est102:BG562602	+	128.50	299.26	1.5e-07	697	! BG562602 602581372F1 NIH_MGC_76
gb_est102:BG564407	+	128.50	297.67	1.8e-07	822	! BG564407 602584345F1 NIH_MGC_76
gb_est29:AL563977	-	128.50	297.43	1.9e-07	843	! AL563977 AL563977 LTI_NFL001_NE
gb_est42:AW111218	-	128.00	298.66	1.6e-07	653	! AW111218 MT15218 mouse liver, d
gb_est40:AV657144	+	127.50	303.98	8.2e-08	331	! AV657144 AV657144 GLC Homo sapi
gb_est40:AV658016	+	127.50	303.41	8.8e-08	351	! AV658016 AV658016 GLC Homo sapi
gb_est40:AV659081	+	127.50	303.17	9.1e-08	360	! AV659081 AV659081 GLC Homo sapi
gb_est40:AV682237	+	127.50	303.11	9.2e-08	362	! AV682237 AV682237 GKB Homo sapi
gb_est40:AV654397	+	127.50	302.22	1.0e-07	397	! AV654397 AV654397 GLC Homo sapi
gb_est105:H72811	+	127.50	302.10	1.0e-07	402	! H72811 yu07f10.r1 Soares fetal
gb_est40:AV645675	+	127.50	301.94	1.1e-07	409	! AV645675 AV645675 GLC Homo sapi
gb_est105:H93888	+	127.50	301.52	1.1e-07	427	! H93888 yv08g08.r1 Soares fetal
gb_est40:AV683948	+	127.50	298.14	1.7e-07	606	! AV683948 AV683948 GKC Homo sapi
gb_est40:AV654309	+	127.50	298.13	1.7e-07	607	! AV654309 AV654309 GLC Homo sapi
gb_est40:AV655831	+	127.50	298.06	1.7e-07	611	! AV655831 AV655831 GLC Homo sapi
gb_est40:AV645552	+	127.50	297.35	1.9e-07	658	! AV645552 AV645552 GLC Homo sapi
gb_est40:AV683159	+	127.50	296.02	2.3e-07	755	! AV683159 AV683159 GKC Homo sapi
gb_est15:AI065103	+	127.00	298.65	1.6e-07	506	! AI065103 HA0939 Human fetal liv
gb_est19:AI333042	-	127.00	298.29	1.7e-07	525	! AI333042 qq31h04.x1 Soares_NhHM
gb_est19:AI336247	-	127.00	297.28	1.9e-07	583	! AI336247 qt45b07.x1 Soares_feta
gb_est40:AV690405	+	127.00	296.27	2.2e-07	647	! AV690405 AV690405 GKC Homo sapi
gb_est42:AW111217	-	126.50	296.47	2.1e-07	558	! AW111217 MT2416 mouse liver, di
gb_est17:AI246161	-	126.00	296.54	2.1e-07	487	! AI246161 qi29b06.x1 Soares_N
gb_est40:AV682469	+	126.00	293.54	3.1e-07	665	! AV682469 AV682469 GKB Homo s
gb_est17:AI195212	-	126.00	290.42	4.7e-07	918	! AI195212 ui62a04.x1 Sugano m
gb_est40:AV657143	+	125.00	297.59	1.9e-07	338	! AV657143 AV657143 GLC Homo s

seq_name: gb_est45:AW303354

seq_documentation_block:
LOCUS AW303354 443 bp mRNA EST 18-JAN-2000
DEFINITION xv18b01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2813449 3' similar to SW:IBP_CARCR P00993 CHELONIANIN ;, mRNA sequence.
ACCESSION AW303354
VERSION AW303354.1 GI:6713034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2813449"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 137 a 84 c 77 g 145 t
ORIGIN

alignment_scores:
Quality: 419.00 Length: 75
Ratio: 5.587 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-177_COPY_25_99 x AW303354/rev ..

Align seg 1/1 to reverse of: AW303354 from: 1 to: 443

```
      1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    432 GGTGTTAATAAAATTGCGGAGAAGATATGTGGAGACCTCAAAGATCCCTG 383

     17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    382 CAAATTGGACATGAATTTTGGAAGCTGCTATGAAGTTCACTTTAGATATT 333

     34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    332 TCTACAACAGAACCTCCAAAAGATGTGAAACTTTTGTCTTCTCCGGCTGT 283
```

```
     51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    282 AATGGCAACCTTAACAACTTCAAGCTTAAAATAGAACGTGAAGTAGCCTG 233

     67 sValAlaLysTyrLysProProArg 75
        |||||||||||||||||||||||
    232 TGTTGCAAAATACAAACCACCGAGG 208
```

seq_name: gb_htc:AK020315 .

seq_documentation_block:

LOCUS AK020315 611 bp mRNA HTC 08-FEB-2001
DEFINITION Mus musculus adult male epididymis cDNA, RIKEN full-length enriched library, clone:9230105I15, full insert sequence.
ACCESSION AK020315
VERSION AK020315.1 GI:12860868
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male epididymis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:9230105I15.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods Enzymol. 303, 19-44 (1999)
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 611)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.2. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES Location/Qualifiers
source 1. .611
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1912158"
/db_xref="MGD:MGI:1925489"
/clone="9230105I15"
/sex="male"
/tissue_type="epididymis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 24. .503
/note="putative"
/codon_start=1
/protein_id="BAB32069.1"
/db_xref="GI:12860869"
/translation="MKPTKLGFLLGLSLLCSLSPPVLSGVERLANYLCKDYNDPCLLD
VEPGSCYEVHFRFFYNQTAKQCQIFLFTGCNGNLNNFKLKIDCDVTCHETYKNPPLPS
GDKRKRSLRALATKNLATDWRLTTLQPGRLRQTERKETLLQRARRESRRQPLARHHT"
polyA_signal 591. .596
/note="putative"
polyA_site 611
/note="putative"
BASE COUNT 174 a 152 c 145 g 140 t
ORIGIN

alignment_scores:
Quality: 269.00 Length: 74
Ratio: 4.270 Gaps: 0
Percent Similarity: 85.135 Percent Identity: 60.811

alignment_block:
US-09-215-435-177_COPY_25_99 x AK020315 ..

Align seg 1/1 to: AK020315 from: 1 to: 611

```
      1 GlyValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCy 17
        ||||||:::::::::||:::   :::|||   |||   :::||||||||
     96 GGTGTTGAGAGGCTAGCTAACTATCTCTGCAAAGACTACAACGATCCCTG 145

     17 sLysLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrP 34
        |   ||||||::::::   ||||||||||||||||||||||||||||:::|
    146 TCTGTTGGATGTGGAGCCTGGCAGCTGCTATGAAGTACATTTTAGATTTT 195

     34 heTyrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCys 50
        ||||||||:::|||:::|||:::|||:::   |||:::|||:::||||||
    196 TCTACAACCAAACTGCCAAACAGTGTCAGATTTTTCTGTTCACTGGCTGC 245

     51 AsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCy 67
        ||||||||||||||||||||||||||||||||||:::   :::|||:::||
    246 AATGGCAACCTTAATAACTTCAAGCTGAAAATAGACTGTGATGTAACTTG 295

     67 sValAlaLysTyrLysProPro 74
        |      ||||||   |||
    296 TCACGAAACATACAAAAACCCT 317
```

```
seq_name: gb_gss32:AZ763929

seq_documentation_block:
LOCUS       AZ763929      570 bp    DNA             GSS       16-FEB-2001
DEFINITION  1M0559C11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0559C11 R, DNA sequence.
ACCESSION   AZ763929
VERSION     AZ763929.1  GI:12878349
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0559  row: C  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0559C11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      179 a    123 c    126 g    142 t
ORIGIN


alignment_scores:
            Quality:  235.00            Length:     60
              Ratio:  4.519               Gaps:      0
 Percent Similarity:  86.667   Percent Identity: 65.000

alignment_block:
```

```
US-09-215-435-177_COPY_25_99 x AZ763929/rev   ..

Align seg 1/1  to reverse of: AZ763929  from: 1  to: 570

     15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
        |||||||||   ||||||::::::   ||||||||||||||||||||||
    208 GATCCCTGTCTGTTGGATGTGGAGCCTGGCAGCTGCTATGAAGTACATTT 159

     31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
        ||||:::|||||||||:::|||:::|||:::|||:::   |||:::|||:
    158 TAGATTTTTCTACAACCAAACTGCCAAACAGTGTCAGATTTTTCTGTTCA 109

     48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
        ::|||||||||||||||:::||||||||||||||||||||||::   :::
    108 CTGGCTGCAATGGCAACTTTAATAACTTCAAGCTGAAAATAGACTGTGAT 59

     65 ValAlaCysValAlaLysTyrLysProPro 74
        |||:::|||         ||||||   |||
     58 GTAACTTGTCACGAAACATACAAAAACCCT 29

seq_name: gb_est15:AI015730

seq_documentation_block:
LOCUS       AI015730      496 bp    mRNA            EST       27-AUG-1998
DEFINITION  ot71b08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1622199 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN
            );, mRNA sequence.
ACCESSION   AI015730
VERSION     AI015730.1  GI:3230066
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2157   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1622199"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      105 a    145 c    128 g    118 t
ORIGIN


alignment_scores:
            Quality:  132.00            Length:     54
              Ratio:  3.882               Gaps:      0
 Percent Similarity:  62.963   Percent Identity: 48.148
```

```
alignment_block:
 US-09-215-435-177_COPY_25_99 x AI015730/rev   ..

 Align seg 1/1  to reverse of: AI015730  from: 1  to: 496

      14 LysAspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHi 30
         :::|||   |||:::|||      :::   |||||||||      :::
     493 GAAGATTCCTGCCAGCTGGGCTACTCGGCCGGTTCCTGCATGGGAATGAC 444

      30 sPheArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValP 47
            ||||||||||||||   ||||||      ||||||||||||   :
     443 CAGCAGGTATTTCTATAATGGTACATCCATGGCCTGTGAGACTTTCCAGT 394

      47 heSerGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArg 63
         :::::||||||   ||||||   |||||||||      :::   |||
     393 ACGGCGGCTGCATGGGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGT 344

      64 GluValAlaCys 67
            :::|||
     343 CTGCAGACCTGC 332

seq_name: gb_est18:AI256949

seq_documentation_block:
LOCUS       AI256949      552 bp    mRNA            EST       17-NOV-1998
DEFINITION  LP04037.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP04037 5prime, mRNA sequence.
ACCESSION   AI256949
VERSION     AI256949.1  GI:3864474
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Sequence is the complete cDNA Insert. hit genomic sequence DS00216;
            hit genomic sequence DS02911; hit genomic sequence DS03364
            Plate: 40  row: D  column: 1
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LP04037"
                     /clone_lib="LP Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /lab_host="DH5-alpha"
                     /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                     Site_2: XhoI; Sized fractionated cDNAs were directly
                     ligated into pOT2. Plasmid cDNA library.
                     "
BASE COUNT      154 a    124 c    122 g    152 t
ORIGIN


alignment_scores:
            Quality:  131.00            Length:     56
              Ratio:  3.639               Gaps:      0
 Percent Similarity:  64.286   Percent Identity: 41.071
```

```
alignment_block:
 US-09-215-435-177_COPY_25_99 x AI256949   ..

 Align seg 1/1  to: AI256949  from: 1  to: 552

      15 AspProCysLysLeuAspMetAsnPheGlySerCysTyrGluValHisPh 31
         :::   |||            :::   |||   |||:::   :::   ::
     163 GAGGACTGTCATCAGCCCAAGGAAACCGGTCGCTGTTTCGCTCTGTTCTA 212

      31 eArgTyrPheTyrAsnArgThrSerLysArgCysGluThrPheValPheS 48
         :|||||   ||||||      ::::::   ||||||   |||||||:::
     213 TCGCTACGCCTACAACGTGGATACACAATCCTGCGAGGAGTTCGTTTACG 262

      48 erGlyCysAsnGlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGlu 64
         ::|||||   ||||||   |||||||||:::   |||   :::   |||
     263 GTGGATGTGCCGGCAACAAGAACAACTTCGAATCCAAGGAGCAGTGCGAA 312

      65 ValAlaCysValAlaLys 70
            ||||||::::::|||
     313 CAGGCATGTTTGGTAAAG 330

seq_name: gb_est104:D83828

seq_documentation_block:
LOCUS       D83828        552 bp    mRNA            EST       15-APR-1996
DEFINITION  HUMHBC4837 Human pancreatic islet Homo sapiens cDNA similar to pp5,
            mRNA sequence.
ACCESSION   D83828
VERSION     D83828.1  GI:1262515
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Takeda,J.
  TITLE     Human pancreatic islet ESTs
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Jun Takeda
            Institute for Molecular and Cellular Regulation, Gunma University
            3-39-15 Showa-machi, Maebashi Gunma 371, Japan
            Tel: 272-20-8856
            Fax: 272-20-8896
            Email: jtakeda@sb.gunma-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human pancreatic islet"
                     /note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
                     I; mRNA was prepared from normal adult human islets. cDNA
                     was directionally synthesized from the Xho I in the vector
                     to the EcoRI site. cDNA was size fractionated to remove
                     sequences <1000 bp in size."
BASE COUNT      146 a    129 c    136 g    135 t      6 others
ORIGIN


alignment_scores:
            Quality:  129.00            Length:     70
              Ratio:  3.000               Gaps:      0
 Percent Similarity:  61.429   Percent Identity: 34.286

alignment_block:
 US-09-215-435-177_COPY_25_99 x D83828   ..

 Align seg 1/1  to: D83828  from: 1  to: 552

       6 AlaGluLysIleCysGlyAspLeuLysAspProCysLysLeuAspMetAs 22
         |||::::::      |||:::      :::   |||   |||   :::||
       1 GCTCAGGAGCCAACAGGAAATAACGCGGAGATCTGTCTCCTGCCCCTAAA 50

      22 nPheGlySerCysTyrGluValHisPheArgTyrPheTyrAsnArgThrS 39
```

```
      |:::|||   |||      :::   :::||||||:::|||:::|||   :
     51 CTACGGACCCTGCCGGGCCCTACTTCTCCGTTACTACTACGACAGGTACA 100

     39 erLysArgCysGluThrPheValPheSerGlyCysAsnGlyAsnLeuAsn 55
        :::::   |||:::   |||:::::::::|||||||::|||||||   |||
    101 CGCAGAGCTGCCGCCAGTTCCTGTACGGGGGCTGCGAGGGCAACGCCAAC 150

     56 AsnPheLysLeuLysIleGluArgGluValAlaCysValAlaLysTyrLy 72
        ||||||                   :::   ||||||           ||
    151 AATTTCTACACCTGGGAGGCTTGCGACNATGCTTGCTGGAGGATANAAAA 200

     72 sProProArg 75
        |   |||:::
    201 AGTTCCCAAA 210

seq_name: gb_est109:T74655

seq_documentation_block:
LOCUS       T74655          350 bp    mRNA            EST       02-MAR-1995
DEFINITION  yc57e11.r1 Stratagene liver (#937224) Homo sapiens cDNA clone
            IMAGE:84812 5' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);,
            mRNA sequence.
ACCESSION   T74655
VERSION     T74655.1  GI:691330
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 62
            High qality sequence stops: 281 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 62   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 281.
FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Homo sapiens"
                     /db_xref="GDB:501869"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:84812"
                     /clone_lib="Stratagene liver (#937224)"
                     /sex="male"
                     /dev_stage="49 years old"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Hepatectomy from normal male caucasian. Average insert
                     size: 1.1 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       78 a     92 c    103 g     76 t      1 others
ORIGIN


alignment_scores:
            Quality:  128.50            Length:     66
              Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x T74655   ..

 Align seg 1/1  to: T74655  from: 1  to: 350

      2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
        |||:::::::::::::::|||              :::|||   |||::
     14 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 48

     18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
        :|||      :::   |||   |||      :::      ||||||||||
     49 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 98

     35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
        |||||   ||||||      ||||||||||||   ::::::||||||
     99 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 148

     52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
        ||||||   |||||||||      :::   |||         :::|||
    149 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 196

seq_name: gb_est7:AA424642

seq_documentation_block:
LOCUS       AA424642        413 bp    mRNA            EST       16-OCT-1997
DEFINITION  zv91d02.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:767139 5'
            similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN);, mRNA sequence.
ACCESSION   AA424642
VERSION     AA424642.1  GI:2103595
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:767139"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
```

```
                    340488-345479, and 484488-489479."
BASE COUNT       93 a    114 c    126 g     80 t
ORIGIN


alignment_scores:
            Quality:  128.50              Length:     66
              Ratio:  3.212                 Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AA424642   ..

 Align seg 1/1  to: AA424642  from: 1  to: 413

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||                ::::|||   |||::
     100 GTAACTGAAGTCACCAAGAAA..............GAAGATTCCTGCCA 134

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     135 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 184

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     185 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 234

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     235 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 282

seq_name: gb_est14:AF063548

seq_documentation_block:
LOCUS       AF063548      511 bp    mRNA            EST       16-JUN-1999
DEFINITION  AF063548 Homo sapiens library (Yu Y) Homo sapiens cDNA clone HA0104
            , mRNA sequence.
ACCESSION   AF063548
VERSION     AF063548.1  GI:5080918
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Yu,Y., Dong,C. and He,F.
  TITLE     Finding some ESTs
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Yongtao Yu
            Department of Hematology
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-68159479
            Fax: 0086-10-68214653
            Email: yyt48@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HA0104"
                     /clone_lib="Homo sapiens library (Yu Y)"
BASE COUNT      114 a    140 c    155 g    102 t
ORIGIN


alignment_scores:
            Quality:  128.50              Length:     66
              Ratio:  3.212                 Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AF063548   ..

Align seg 1/1  to: AF063548  from: 1  to: 511

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||                ::::|||   |||::
     193 GTAACTGAAGTCACCAAGAAA..............GAAGATTCCTGCCA 227

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     228 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 277

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     278 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 327

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     328 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 375

seq_name: gb_est16:AI110674

seq_documentation_block:
LOCUS       AI110674      511 bp    mRNA            EST       11-NOV-1999
DEFINITION  HA0104 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI110674
VERSION     AI110674.1  GI:6359539
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Yongtao Yu
            Department of Hematology
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-68159479
            Fax: 0086-10-68214653
            Email: yyt48@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      114 a    140 c    155 g    102 t
ORIGIN


alignment_scores:
            Quality:  128.50              Length:     66
              Ratio:  3.212                 Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AI110674   ..

 Align seg 1/1  to: AI110674  from: 1  to: 511

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||                ::::|||   |||::
     193 GTAACTGAAGTCACCAAGAAA..............GAAGATTCCTGCCA 227

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     228 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 277
```

```
      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     278 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 327

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     328 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 375
```

seq_name: gb_est40:AV720273

seq_documentation_block:

```
LOCUS       AV720273      520 bp    mRNA            EST       16-OCT-2000
DEFINITION  AV720273 GLC Homo sapiens cDNA clone GLCGBA03 5', mRNA sequence.
ACCESSION   AV720273
VERSION     AV720273.1  GI:10817425
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            ,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA GLC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCGBA03"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      107 a    152 c    137 g    124 t
ORIGIN
```

alignment_scores:

```
            Quality:  128.50            Length:     66
              Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909
```

alignment_block:

US-09-215-435-177_COPY_25_99 x AV720273/rev ..

Align seg 1/1 to reverse of: AV720273 from: 1 to: 520

```
       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||             :::|||   |||::
     512 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 478

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     477 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 428

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     427 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 378

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     377 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 330
```

seq_name: gb_est17:AI207711

seq_documentation_block:

```
LOCUS       AI207711      521 bp    mRNA            EST       11-NOV-1999
DEFINITION  HA3093 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI207711
VERSION     AI207711.1  GI:6361726
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Chenggang Zhang
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Email: zhang_chenggang@hotmail.com.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      116 a    142 c    157 g    106 t
ORIGIN
```

alignment_scores:

```
            Quality:  128.50            Length:     66
              Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909
```

alignment_block:

US-09-215-435-177_COPY_25_99 x AI207711 ..

Align seg 1/1 to: AI207711 from: 1 to: 521

```
       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||             :::|||   |||::
     193 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 227

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     228 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 277

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     278 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 327

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     328 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 375
```

seq_name: gb_est17:AI207590

seq_documentation_block:

```
LOCUS       AI207590      571 bp    mRNA            EST       11-NOV-1999
DEFINITION  HA2954 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI207590
VERSION     AI207590.1  GI:6361600
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Chenggang Zhang
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Email: zhang_chenggang@hotmail.com.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      113 a    174 c    149 g    135 t
ORIGIN


alignment_scores:
            Quality:  128.50            Length:     66
              Ratio:  3.212               Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AI207590/rev   ..

 Align seg 1/1  to reverse of: AI207590  from: 1  to: 571

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||               :::|||   |||::
     511 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 477

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||      :::      ||||||||||
     476 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 427

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     426 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 377

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     376 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 329

 seq_name: gb_est40:AV658823

 seq_documentation_block:
 LOCUS       AV658823      626 bp    mRNA            EST       07-SEP-2000
 DEFINITION  AV658823 GLC Homo sapiens cDNA clone GLCFQE08 3', mRNA sequence.
 ACCESSION   AV658823
 VERSION     AV658823.1  GI:9879837
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 626)
   AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
             , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
             Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
             ,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
   TITLE     Homo sapiens cDNA clone
   JOURNAL   Unpublished (2000)
 COMMENT     Contact: Zeguang Han
             Chinese National Human Genome Center at Shanghai
             351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
             201203, P. R. China
             Tel: 86-21-50801919(ex.45)
             Fax: 86-21-50801922
             Email: hanzg@chgc.sh.cn
             This clone is available at CHGC in Shanghai.
 FEATURES             Location/Qualifiers
      source          1. .626
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="GLCFQE08"
                      /clone_lib="GLC"
                      /tissue_type="corresponding non cancerous liver tissue"
                      /dev_stage="Adult"
                      /lab_host="SOLR"
                      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                      XhoI"
 BASE COUNT      163 a    157 c    183 g    120 t      3 others
 ORIGIN


 alignment_scores:
             Quality:  128.50            Length:     66
               Ratio:  3.212               Gaps:      1
  Percent Similarity:  60.606   Percent Identity: 40.909

 alignment_block:
  US-09-215-435-177_COPY_25_99 x AV658823   ..

  Align seg 1/1  to: AV658823  from: 1  to: 626

        2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
          |||:::::::::::::::|||               :::|||   |||::
       98 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 132

       18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
          :|||      :::   |||   |||      :::      ||||||||||
      133 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 182

       35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
          |||||   ||||||      ||||||||||||   ::::::||||||
      183 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 232

       52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
          ||||||   |||||||||      :::   |||         :::|||
      233 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 280

 seq_name: gb_est25:AI807643

 seq_documentation_block:
 LOCUS       AI807643      644 bp    mRNA            EST       19-DEC-1999
 DEFINITION  wf49d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
             IMAGE:2358919 3' similar to gb:X04225 ALPHA-1-MICROGLOBULIN (HUMAN
             );, mRNA sequence.
 ACCESSION   AI807643
 VERSION     AI807643.1  GI:5394209
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 644)
   AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
   TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
   JOURNAL   Unpublished (1997)
 COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 1328   Std Error: 0.00
             Seq primer: -40UP from Gibco
```

```
           High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2358919"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      128 a    195 c    173 g    148 t
ORIGIN


alignment_scores:
            Quality:  128.50             Length:     66
              Ratio:  3.212                Gaps:      1
 Percent Similarity:  60.606   Percent Identity: 40.909

alignment_block:
 US-09-215-435-177_COPY_25_99 x AI807643/rev   ..

 Align seg 1/1  to reverse of: AI807643  from: 1  to: 644

       2 ValAsnLysIleAlaGluLysIleCysGlyAspLeuLysAspProCysLy 18
         |||:::::::::::::::|||               :::|||   |||::
     508 GTAACTGAAGTCACCAAGAAA.............GAAGATTCCTGCCA 474

      18 sLeuAspMetAsnPheGlySerCysTyrGluValHisPheArgTyrPheT 35
         :|||      :::   |||   |||     :::      ||||||||||
     473 GCTGGGCTACTCGGCCGGTCCCTGCATGGGAATGACCAGCAGGTATTTCT 424

      35 yrAsnArgThrSerLysArgCysGluThrPheValPheSerGlyCysAsn 51
         |||||   ||||||      ||||||||||||   ::::::||||||
     423 ATAATGGTACATCCATGGCCTGTGAGACTTTCCAGTACGGCGGCTGCATG 374

      52 GlyAsnLeuAsnAsnPheLysLeuLysIleGluArgGluValAlaCys 67
         ||||||   |||||||||      :::   |||         :::|||
     373 GGCAACGGTAACAACTTCGTCACAGAAAAGGAGTGTCTGCAGACCTGC 326
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-179 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:43 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pr6:AL157823         +  346.00  614.04   5.8e-26   127384  ! AL157823 Human DNA sequence
gb_htg22:AL512371       +  346.00  612.09   7.4e-26   159482  ! AL512371 Homo sapiens chromc
gb_htg23:AL590400       +  346.00  611.59   7.9e-26   168884  ! AL590400 Homo sapiens chromc
gb_om:CCCOLIP           +  147.00  277.30   3.3e-07      452  ! X53564 Canine mRNA for colipase
gb_om:RABCOLIPAS        +  141.00  265.54   1.5e-06      459  ! L06329 Oryctolagus cuniculus pa
gb_pr10:HUMCOLIP        +  135.50  253.74   6.8e-06      523  ! J02883 Human colipase mRNA, com
gb_pat1:AX014220        +  135.50  251.77   8.7e-06      656  ! AX014220 Sequence 16 from Paten
gb_ro1:MCCOLIPAS        +  134.50  252.36   8.1e-06      490  ! X82998 M.coypus mRNA for colipa
gb_om:ECCOLIPB          +  134.00  251.27   9.3e-06      497  ! X74344 E.caballus mRNA for coli
gb_om:AF148567          +  133.00  249.11   1.2e-05      510  ! AF148567 Sus scrofa pancreatic
gb_om:ECCO1IPA          +  131.00  245.90   1.8e-05      472  ! X74503 E.caballus mRNA for coli
gb_ro2:RATCOLQ          +  122.50  229.06    0.0002      492  ! M58370 Rat colipase mRNA, compl
gb_ro2:RATCOLIP         +  122.00  228.95    0.0002      446  ! M33333 Rat colipase mRNA, compl
gb_in3:CEW02A2          -   96.50  141.35     12.31      35879  ! Z82286 Caenorhabditis elegans
gb_htg24:CEY62E10       -   96.50  124.31     109.37     254217  ! AL031580 Caenorhabditis eleg
gb_htg24:CEY67H2        -   96.50  122.03     146.68     330724  ! AL022475 Caenorhabditis eleg
gb_in3:CEW01B6          +   93.50  137.65     19.78      28122  ! Z68301 Caenorhabditis elegans
gb_htg19:AF207953       +   93.00  124.55     106.05     113351  ! AF207953 Homo sapiens chromc
gb_in2:AF220362         -   90.50  162.88     0.7773      792  ! AF220362 Drosophila melanogaste
gb_htg4:AC013177        +   90.50  116.31     305.47     167596  ! AC013177 Drosophila melanoga
gb_in1:AC007257         +   90.50  115.54     336.96     183007  ! AC007257 Drosophila melanoga
gb_in1:AE003625         -   90.50  112.47     499.90     260624  ! AE003625 Drosophila melanoga
gb_htg20:AL159994       -   89.00  112.30     510.44     190081  ! AL159994 Homo sapiens chromc
gb_htg25:HS183I6        +   89.00  111.97     532.64     197475  ! AL449983 Homo sapiens chromc
gb_pr9:HSJ806M20        -   88.50  114.87     367.26     126586  ! AL132655 Human DNA sequence
gb_htg16:AC079226       -   88.50  111.74     548.52     181364  ! AC079226 Homo sapiens chromc
gb_ov:BSU43200          -   88.00  149.68     4.23        2070  ! U43200 Boreogadus saida antifr
gb_htg23:AP000562       -   88.00  112.62     489.83     146589  ! AP000562 Homo sapiens chromc
gb_htg24:AP002351       -   88.00  111.52     564.49     166465  ! AP002351 Homo sapiens chromc
gb_pr6:AL355593         -   88.00  111.44     570.54     168065  ! AL355593 Human DNA sequence
gb_htg24:AP002761       +   88.00  110.30     659.80     191451  ! AP002761 Homo sapiens chromc
gb_pl3:F21F23           -   87.50  117.03     278.52      79054  ! AC027656 Sequence of BAC F21F
gb_pr7:AP000885         -   87.50  116.22     308.82      86719  !.AP000885 Homo sapiens genomic
gb_pr7:AP001708         -   87.50  104.34     1.4e+03     340000  ! AP001708 Homo sapiens genomi
gb_pr10:HSY10344        -   87.00  145.32     7.39        2734  ! Y10344 H.sapiens GalNAc-T2 gen
gb_htg19:AL136988       +   87.00  110.68     628.91     146755  ! AL136988 Homo sapiens chromc
gb_pr6:AL360088         +   87.00  109.71     711.96     164013  ! AL360088 Human DNA sequence
gb_htg10:AC025391       +   87.00  109.69     713.25     164280  ! AC025391 Homo sapiens chromc
gb_htg9:AC024008        +   87.00  109.69     713.60     164352  ! AC024008 Homo sapiens chromc
gb_htg11:AC027244       +   87.00  109.61     721.33     165947  ! AC027244 Homo sapiens chromc
gb_htg8:AC022621        -   87.00  109.55     726.97     167110  ! AC022621 Homo sapiens clone
gb_pr3:AC011475         -   87.00  108.90     789.57     179953  ! AC011475 Homo sapiens chr
gb_htg20:AL158168       +   87.00  108.70     810.55     184233  ! AL158168 Homo sapiens chr
gb_htg13:AC067715       +   87.00  108.34     848.90     192027  ! AC067715 Mus musculus chr
gb_pr8:HS69B13          -   87.00  107.97     889.88     200317  ! AL035698 Human DNA sequen

seq_name: gb_pr6:AL157823

seq_documentation_block:
LOCUS       AL157823   127384 bp    DNA             PRI       10-FEB-2001
DEFINITION  Human DNA sequence from clone RP3-51008 on chromosome 6, complete
            sequence.
ACCESSION   AL157823
VERSION     AL157823.9  GI:12750800
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127384)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 12, 2001 this sequence version replaced gi:12709869.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-51008 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-51008 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-51008 is at 1 in this sequence. The
            true left end of clone RP3-422H11 is at 127285 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .127384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-51008"
                     /clone_lib="RPCI-3"
     misc_feature    6404
                     /note="Weak data in alu."
     misc_feature    complement(95875. .96037)
                     /note="match: STS: Em:G13580"
     misc_feature    95894. .96129
                     /note="match: STS: Em:G14863"
BASE COUNT    33605 a  30130 c  31408 g  32241 t
ORIGIN
```

```
alignment_scores:
           Quality:  346.00              Length:    349
             Ratio:  3.681                Gaps:      2
 Percent Similarity:  26.934   Percent Identity: 26.074

alignment_block:
 US-09-215-435-179 x AL157823   ..

 Align seg 1/1  to: AL157823  from: 1  to: 127384

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   ||||||:::           :::|||   ||||||||||||||
   87854 CGGTGGAGCCTGGCGGTGCCGTGTCCCTCCCTGCAGGAGCTCAAGGAGTC 87903

      38 rCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProA 55
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   87904 TTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAG 87953

      55 spAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCys 71
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   87954 ACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGT 88003

      72 GlnThrGlnVal..Phe..................................... 76
         ||||||||||||  :::
   88004 CAAACGCAGGTGGGTATCGCCGCCCGGGGGGAGCCAGAGGGGATCCAGGG 88053

      76 ..................................................... 76

   88054 GAAGTGGGAGCCAGGGCGGGCCCAGATTCCTGGGGAGGAAAGAAGGGTAG 88103

      76 ..................................................... 76

   88104 GTGGGCGGAAATGCCCTGGAGCCTGGAATTCCCCCGTGGGACTGCACCTC 88153

      76 ..................................................... 76

   88154 CAAGCCTTCTGGGGGAGCCTGTGGACGGAACCAATATGGTAGCCCACCCT 88203

      76 ..................................................... 76

   88204 AGGGTTTTGGAATCCTGGATTGGGGCGGGAAGGCGATGTTTCACTTTCTA 88253

      76 ..................................................... 76

   88254 ATGTGGGTTCACATTCTCCAGGAGTCTTCTCAGCAAATTCCGAAAGTAGT 88303

      76 ..................................................... 76

   88304 ACAGTAAATGGGCACAGCTAGGTCCAGCCTCTGCCCCAGAGCCCCGTCTT 88353

      76 ..................................................... 76

   88354 CATCCAGCTCAGGAGTTCTGCTGTGGCCCCGGGTGTTCCTGGTGGAGGAT 88403

      76 ..................................................... 76

   88404 GGGCAGAGGCAGCCTCTGAGCTGGGTGTGTAGGGGAGGGAGTTGATTTGC 88453

      76 ..................................................... 76

   88454 CCATGGAAGGATTTAAAGACAAGACATAGAGGTCCTGCGATGGTGTGTCC 88503

      76 ..................................................... 76

   88504 ATGGACTTAGGACAGGCAGATGGGTCAGTGACAGGCAGGGTTACAAGGTC 88553

      76 ..................................................... 76

   88554 TGGAGCTGGGAGGACACTGAACAATCCAGGAGCCAGCCTGGAGAAGGCAG 88603

      76 ..................................................... 76

   88604 CCAGAGGTGCCACTCTGAGCAGCTGGCAAGCTATCATTTATCCCTGGATC 88653

      76 ..................................................... 76

   88654 CTGGCTGTGGGCAAGCACATGGGCCCTGGAGCTGGGCTGGGCTGGGCTTA 88703

      76 ..................................................... 76

   88704 GGGGCTCAGGGAAGAGCTGCCGCCAAGGTCGAGGTCAAAGTCCTGTCTTT 88753

      77 ..........PheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu 89
                  |||||||||||||||||||||||||||||||||||||||||
   88754 CCCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTG 88803

      90 ThrCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyAr 106
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   88804 ACTTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCG 88853

     106 gCysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
         |||||||||||||||||||||||||||||||||||||||||||||
   88854 TTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 88899

seq_name: gb_htg22:AL512371

seq_documentation_block:
LOCUS       AL512371   159482 bp    DNA             HTG       21-MAR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-238F24, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL512371
VERSION     AL512371.4  GI:13274864
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159482)
  AUTHORS   Clark,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 12, 2001 this sequence version replaced gi:12964529.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA238F24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 154970 bases at least Q40
            Consensus quality: 156246 bases at least Q30
            Consensus quality: 156738 bases at least Q20
            Insert size: 159282; sum-of-contigs
            Insert size: 165487; 7.5% error; agarose-fp
            Quality coverage: 6.35x in Q20 bases; sum-of-contigs Quality
            coverage: 6.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     3066: contig of 3066 bp in length
            *   3067  3166: gap of      100 bp
            *   3167   155002: contig of 151836 bp in length
            * 155003 155102: gap of      100 bp
            * 155103   159482: contig of 4380 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .159482
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-238F24"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .3066
                     /note="assembly_fragment:02384
                     fragment_chain:1"
     misc_feature    3167. .155002
                     /note="assembly_fragment:02553
                     fragment_chain:1"
     misc_feature    155103. .159482
                     /note="assembly_fragment:02886"
BASE COUNT    42530 a  37388 c  38532 g  40831 t    201 others
ORIGIN


alignment_scores:
            Quality:  346.00             Length:    349
              Ratio:  3.681                Gaps:      2
 Percent Similarity:  26.934   Percent Identity: 26.074

alignment_block:
 US-09-215-435-179 x AL512371   ..

 Align seg 1/1  to: AL512371  from: 1  to: 159482

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   ||||||:::           :::|||   ||||||||||||||
  101396 CGGTGGAGCCTGGCGGTGCCGTGTCCCTCCCTGCAGGAGCTCAAGGAGTC 101445

      38 rCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProA 55
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  101446 TTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAG 101495

      55 spAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCys 71
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  101496 ACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGT 101545

      72 GlnThrGlnVal..Phe..................................... 76
         ||||||||||||  :::
  101546 CAAACGCAGGTGGGTATCGCCGCCCGGGGGGAGCCAGAGGGGATCCAGGG 101595

      76 ..................................................... 76

  101596 GAAGTGGGAGCCAGGGAGGGCCCAGATTCCTGGGGAGGAAAGAAGGGTAG 101645

      76 ..................................................... 76

  101646 GTGGGCGGAAATGCCCTGGAGCCTGGAATTCCCCCGTGGGACTGCACCTC 101695

      76 ..................................................... 76

  101696 CAAGCCTTCTGGGGGAGCCTGTGGACGGAACCAATATGGTAGCCCACCCT 101745

      76 ..................................................... 76

  101746 AGGGTTTTGGAATCCTGGATTGGGGCGGGAAGGCGATGTTTCACTTTCTA 101795

      76 ..................................................... 76

  101796 ATGTGGGTTCACATTCTCCAGGAGTCTTCTCAGCAAATTCCGAAAGTAGT 101845

      76 ..................................................... 76

  101846 ACAGTAAATGGGCACAGCTAGGTCCAGCCTCTGCCCCAGAGCCCCGTCTT 101895

      76 ..................................................... 76

  101896 CATCCAGCTCAGGAGTTCTGCTGTGGCCCCGGGTGTTCCTGGTGGAGGAT 101945

      76 ..................................................... 76

  101946 GGGCAGAGGCAGCCTCTGAGCTGGGTGTGTAGGGGAGGGAGTTGATTTGC 101995

      76 ..................................................... 76

  101996 CCATGGAAGGATTTAAAGACAAGACATAGAGGTCCTGCGATGGTGTGTCC 102045

      76 ..................................................... 76

  102046 ATGGACTTAGGACAGGCAGATGGGTCAGTGACAGGCAGGGTTACAAGGTC 102095

      76 ..................................................... 76

  102096 TGGAGCTGGGAGGACACTGAACAATCCAGGAGCCAGCCTGGAGAAGGCAG 102145

      76 ..................................................... 76

  102146 CCAGAGGTGCCACTCTGAGCAGCTGGCAAGCTATCATTTATCCCTGGATC 102195

      76 ..................................................... 76

  102196 CTGGCTGTGGGCAAGCACATGGGCCCTGGAGCTGGGCTGGGCTGGGCTTA 102245

      76 ..................................................... 76

  102246 GGGGCTCAGGGAAGAGCTGCCGCCAAGGTCGAGGTCAAAGTCCTGTCTTT 102295

      77 ..........PheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu 89
                  ||||||||||||||||||||||||||||||||||||||||
  102296 CCCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTG 102345

      90 ThrCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyAr 106
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  102346 ACTTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCG 102395

     106 gCysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
         |||||||||||||||||||||||||||||||||||||||||||||
  102396 TTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 102441

seq_name: gb_htg23:AL590400

seq_documentation_block:
LOCUS       AL590400   168884 bp    DNA             HTG       08-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-569I12, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   AL590400
VERSION     AL590400.2  GI:13568356
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168884)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:13559131.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA569I12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 164511 bases at least Q40
            Consensus quality: 165952 bases at least Q30
```

```
          Consensus quality: 166791 bases at least Q20
          Insert size: 167584; sum-of-contigs
          Insert size: 185239; 2.7% error; agarose-fp
          Quality coverage: 6.19x in Q20 bases; sum-of-contigs Quality
          coverage: 5.74x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     2920: contig of 2920 bp in length
          *     2921 3020: gap of      100 bp
          *     3021    32168: contig of 29148 bp in length
          *    32169 32268: gap of      100 bp
          *    32269    36080: contig of 3812 bp in length
          *    36081 36180: gap of      100 bp
          *    36181    45704: contig of 9524 bp in length
          *    45705 45804: gap of      100 bp
          *    45805    59680: contig of 13876 bp in length
          *    59681 59780: gap of      100 bp
          *    59781    61890: contig of 2110 bp in length
          *    61891 61990: gap of      100 bp
          *    61991    72185: contig of 10195 bp in length
          *    72186 72285: gap of      100 bp
          *    72286    76243: contig of 3958 bp in length
          *    76244 76343: gap of      100 bp
          *    76344    81840: contig of 5497 bp in length
          *    81841 81940: gap of      100 bp
          *    81941    86241: contig of 4301 bp in length
          *    86242 86341: gap of      100 bp
          *    86342   129124: contig of 42783 bp in length
          *   129125 129224: gap of      100 bp
          *   129225   131985: contig of 2761 bp in length
          *   131986 132085: gap of      100 bp
          *   132086   151439: contig of 19354 bp in length
          *   151440 151539: gap of      100 bp
          *   151540   168884: contig of 17345 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-569I12"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .2920
                     /note="assembly_fragment:00514
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3021. .32168
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    32269. .36080
                     /note="assembly_fragment:00777
                     fragment_chain:1"
     misc_feature    36181. .45704
                     /note="assembly_fragment:00971
                     fragment_chain:1"
     misc_feature    45805. .59680
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    59781. .61890
                     /note="assembly_fragment:02150
                     fragment_chain:2"
     misc_feature    61991. .72185
                     /note="assembly_fragment:02579
                     fragment_chain:2"
     misc_feature    72286. .76243
                     /note="assembly_fragment:00659
                     fragment_chain:2"
     misc_feature    76344. .81840
                     /note="assembly_fragment:00425
                     fragment_chain:2"
     misc_feature    81941. .86241
                     /note="assembly_fragment:02764
                     fragment_chain:2"
     misc_feature    86342. .129124
                     /note="assembly_fragment:00846
                     fragment_chain:2"
     misc_feature    129225. .131985
                     /note="assembly_fragment:01963"
     misc_feature    132086. .151439
                     /note="assembly_fragment:02120
                     fragment_chain:3"
     misc_feature    151540. .168884
                     /note="assembly_fragment:01758
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
BASE COUNT    46011 a  38577 c  39446 g  43544 t   1306 others
ORIGIN


alignment_scores:
            Quality:  346.00          Length:    349
              Ratio:  3.681             Gaps:      2
 Percent Similarity:  26.934   Percent Identity: 26.074

alignment_block:
 US-09-215-435-179 x AL590400   ..

 Align seg 1/1  to: AL590400  from: 1  to: 168884

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   ||||||:::           :::|||   |||||||||||||||
  114892 CGGTGGAGCCTGGCGGTGCCGTGTCCCTCCCTGCAGGAGCTCAAGGAGTC 114941

      38 rCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProA 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  114942 TTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAG 114991

      55 spAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCys 71
         |||||||||||||||||||||||||||||||||||||||||||||||||||
  114992 ACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGT 115041

      72 GlnThrGlnVal..Phe..................................... 76
         ||||||||||||  :::
  115042 CAAACGCAGGTGGGTATCGCCGCCCGGGGGGAGCCAGAGGGGATCCAGGG 115091

      76 ..................................................... 76

  115092 GAAGTGGGAGCCAGGGAGGGCCCAGATTCCTGGGGAGGAAAGAAGGGTAG 115141

      76 ..................................................... 76

  115142 GTGGGCGGAAATGCCCTGGAGCCTGGAATTCCCCCGTGGGACTGCACCTC 115191

      76 ..................................................... 76

  115192 CAAGCCTTCTGGGGGAGCCTGTGGACGGAACCAATATGGTAGCCCACCCT 115241

      76 ..................................................... 76

  115242 AGGGTTTTGGAATCCTGGATTGGGGCGGGAAGGCGATGTTTCACTTTCTA 115291

      76 ..................................................... 76

  115292 ATGTGGGTTCACATTCTCCAGGAGTCTTCTCAGCAAATTCCGAAAGTAGT 115341

      76 ..................................................... 76

  115342 ACAGTAAATGGGCACAGCTAGGTCCAGCCTCTGCCCCAGAGCCCCGTCTT 115391
```

```
     76 ..................................................  76

 115392 CATCCAGCTCAGGAGTTCTGCTGTGGCCCCGGGTGTTCCTGGTGGAGGAT 115441

     76 ..................................................  76

 115442 GGGCAGAGGCAGCCTCTGAGCTGGGTGTGTAGGGGAGGGAGTTGATTTGC 115491

     76 ..................................................  76

 115492 CCATGGAAGGATTTAAAGACAAGACATAGAGGTCCTGCGATGGTGTGTCC 115541

     76 ..................................................  76

 115542 ATGGACTTAGGACAGGCAGATGGGTCAGTGACAGGCAGGGTTACAAGGTC 115591

     76 ..................................................  76

 115592 TGGAGCTGGGAGGACACTGAACAATCCAGGAGCCAGCCTGGAGAAGGCAG 115641

     76 ..................................................  76

 115642 CCAGAGGTGCCACTCTGAGCAGCTGGCAAGCTATCATTTATCCCTGGATC 115691

     76 ..................................................  76

 115692 CTGGCTGTGGGCAAGCACATGGGCCCTGGAGCTGGGCTGGGCTGGGCTTA 115741

     76 ..................................................  76

 115742 GGGGCTCAGGGAAGAGCTGCCGCCAAGGTCGAGGTCAAAGTCCTGTCTTT 115791

     77 ..........PheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu  89
                 ||||||||||||||||||||||||||||||||||||||||||
 115792 CCCAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTG 115841

     90 ThrCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyAr 106
        ||||||||||||||||||||||||||||||||||||||||||||||||||
 115842 ACTTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCG 115891

    106 gCysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
        |||||||||||||||||||||||||||||||||||||||||||||
 115892 TTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 115937

seq_name: gb_om:CCCOLIP

seq_documentation_block:
LOCUS       CCCOLIP       452 bp    mRNA            MAM       31-MAR-1995
DEFINITION  Canine mRNA for colipase.
ACCESSION   X53564
VERSION     X53564.1  GI:841
KEYWORDS    colipase.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Fukuoka,S.I. and Scheele,G.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1990) Fukuoka S.-I., Scheele G.A., The Charles A.
            Dana Research Institute, Harvard Medical School, Cell and Molecular
            Biology, Dana 533, Beth Israel Hospital, 330 Brookline Ave.,
            Boston, MA 02215, USA
REFERENCE   2  (bases 1 to 452)
  AUTHORS   Fukuoka,S., Taniguchi,Y., Kitagawa,Y. and Scheele,G.
  TITLE     Full length cDNA sequence encoding canine pancreatic colipase
  JOURNAL   Nucleic Acids Res. 18 (18), 5549 (1990)
  MEDLINE   91016846
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /dev_stage="adult"
                     /tissue_type="pancreas"
                     /clone_lib="pUC-9"
                     /clone="pCL1"
     sig_peptide     1. .51
                     /product="signal peptide"
     CDS             1. .339
                     /codon_start=1
                     /product="precursor peptide"
                     /protein_id="CAA37636.1"
                     /db_xref="GI:842"
                     /db_xref="SWISS-PROT:P19090"
                     /translation="MEKILVLLLVALAVVYAVPDPRGIIIHLEDGELCLNSVQCKSKC
                     CHRATGLSLARCAPKASENSECSAKTLYGVYYKCPCERGLTCEGDKSIVGSITNTNFG
                     VCHDAGRSKK"
     mat_peptide     52. .336
                     /product="colipase"
     misc_feature    429. .435
                     /note="polyA signal"
     polyA_site      452
                     /note="polyA site"
BASE COUNT       90 a    146 c    117 g     99 t
ORIGIN


alignment_scores:
           Quality:  147.00            Length:     84
             Ratio:  2.625               Gaps:      1
 Percent Similarity:  66.667   Percent Identity: 36.905

alignment_block:
 US-09-215-435-179 x CCCOLIP   ..

 Align seg 1/1  to: CCCOLIP  from: 1  to: 452

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         ||||||      |||   |||:::::::::   :::|||::::::   ||
      82 CTGGAAGACGGCGAGCTCTGTCTGAACAGCGTCCAGTGCAAGAGCAAGTG 131

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         ||||:::||||||               ::::::||||||   |||:::|
     132 CTGCCACCGGGCCACCGGGCTGAGCCTGGCCCGCTGCGCACCCAAGGCCA 181

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
         |||||:::|||   |||:::::::::::::::::::|||   |||
     182 GCGAGAACAGCGAGTGCTCTGCCAAGACGCTCTATGGGGTTTACTACAAG 231

      83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn......Gl 97
         |||||||||   |||:::|||||||||      :::|||:::
     232 TGTCCCTGTGAGCGGGGCCTGACCTGTGAGGGTGACAAGTCCATCGTGGG 281

      97 uLysTrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnL 114
          :::      :::      :::|||   |||:::      ||||||:::|
     282 CTCCATCACCAATACCAACTTTGGTGTCTGCCACGATGCTGGGCGCTCCA 331

     114 ys 114
         ||
     332 AG 333

seq_name: gb_om:RABCOLIPAS

seq_documentation_block:
LOCUS       RABCOLIPAS    459 bp    mRNA            MAM       05-OCT-1993
DEFINITION  Oryctolagus cuniculus pancreatic colipase mRNA, complete cds.
ACCESSION   L06329
VERSION     L06329.1  GI:164893
KEYWORDS    colipase.
SOURCE      Oryctolagus cuniculus male adult pancreas cDNA to mRNA.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Colwell,N.S., Aleman-Gomez,J.A., Sasser,T. and Kumar,V.B.
```

```
  TITLE     Cloning and characterization of rabbit pancreatic colipase
  JOURNAL   Int. J. Biochem. 25 (6), 885-890 (1993)
  MEDLINE   93345715
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
     sig_peptide     39. .89
     CDS             39. .362
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA02911.1"
                     /db_xref="GI:164894"
                     /translation="MEKVLVLLLVALSVAYAAPGPRGIVINLEEGELCLNSAQCKSGC
                     CHHSSALSLARCAPKASENSECSPQTIYGVYYKCPCERGLTCEGDKSIVGSITNTNFG
                     VCLDV"
     mat_peptide     90. .359
                     /product="colipase"
BASE COUNT       84 a    170 c    122 g     83 t
ORIGIN


alignment_scores:
            Quality:  141.00              Length:     59
              Ratio:  3.279                Gaps:      0
 Percent Similarity:  72.881   Percent Identity: 42.373

alignment_block:
 US-09-215-435-179 x RABCOLIPAS   ..

 Align seg 1/1  to: RABCOLIPAS  from: 1  to: 459

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         ||||||      |||   |||:::::::::   :::|||:::::|||||
     120 CTGGAGGAGGGCGAGCTCTGCTTGAACAGCGCCCAGTGCAAGAGCGGCTG 169

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         ||||:::::::::                :::::||||||   |||:::|
     170 CTGCCACCACTCGAGCGCTCTGAGTCTGGCCCGCTGCGCACCCAAGGCCA 219

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
         |||||:::|||   |||:::   |||:::::::::|||   |||
     220 GCGAGAACAGCGAGTGTTCGCCCCAGACCATCTACGGCGTCTACTACAAG 269

      83 CysProCysLeuArgAsnLeuThrCys 91
         |||||||||   |||:::|||||||||
     270 TGTCCCTGTGAGCGCGGCCTGACCTGC 296

 seq_name: gb_pr10:HUMCOLIP

 seq_documentation_block:
 LOCUS       HUMCOLIP      523 bp    mRNA            PRI       01-NOV-1994
 DEFINITION  Human colipase mRNA, complete cds.
 ACCESSION   J02883
 VERSION     J02883.1  GI:180885
 KEYWORDS    cofactor; colipase; triglyceride lipase.
 SOURCE      Human adult pancreas, cDNA to mRNA.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 523)
   AUTHORS   Lowe,M.E., Rosenblum,J.L., McEwen,P. and Strauss,A.W.
   TITLE     Cloning and characterization of the human colipase cDNA
   JOURNAL   Biochemistry 29 (3), 823-828 (1990)
   MEDLINE   90248429
 COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
             by M.E.Lowe, 17-NOV-1989.
 FEATURES             Location/Qualifiers
      source          1. .523
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="6pter-p21.1"
      mRNA            <1. .523
                      /note="CLP mRNA"
      gene            22. .360
                      /gene="CLPS"
      CDS             22. .360
                      /gene="CLPS"
                      /note="colipase precursor"
                      /codon_start=1
                      /db_xref="GDB:G00-127-277"
                      /protein_id="AAA52054.1"
                      /db_xref="GI:180886"
                      /translation="MEKILILLLVALSVAYAAPGPRGIIINLENGELCMNSAQCKSNC
                      CQHSSALGLARCTSMASENSECSVKTLYGIYYKCPCERGLTCEGDKTIVGSITNTNFG
                      ICHDAGRSKQ"
      sig_peptide     23. .73
                      /gene="CLPS"
                      /note="colipase signal peptide"
      mat_peptide     74. .357
                      /gene="CLPS"
                      /note="colipase"
 BASE COUNT      109 a    173 c    128 g    113 t
 ORIGIN


 alignment_scores:
             Quality:  135.50              Length:    111
               Ratio:  1.908                Gaps:      3
  Percent Similarity:  63.964   Percent Identity: 28.829

 alignment_block:
  US-09-215-435-179 x HUMCOLIP   ..

  Align seg 1/1  to: HUMCOLIP  from: 1  to: 523

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::       :::|||
      31 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 80

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
                :::::         ||||||       |||   |||::::::::
      81 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 130

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::::
     131 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 180

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         ::::::|||:::::::   :::|||||||:::|||   |||::::::::::::
     181 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 230

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         :::::::|||   |||      |||||||||   |||:::||||||||||
     231 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 279

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
                :::::::::|||   :::               :::
     280 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 321

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::       ||||||:::|||
     322 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 354

  seq_name: gb_pat1:AX014220

  seq_documentation_block:
  LOCUS       AX014220      656 bp    DNA             PAT       07-SEP-2000
  DEFINITION  Sequence 16 from Patent WO9954446.
  ACCESSION   AX014220
  VERSION     AX014220.1  GI:10040621
  KEYWORDS    .
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
            Pilarsky,C.
  TITLE     Human nucleic acid sequences from normal pancreas tissue
  JOURNAL   Patent: WO 9954446-A 16 28-OCT-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      165 a    200 c    162 g    129 t
ORIGIN


alignment_scores:
            Quality:  135.50             Length:    111
              Ratio:  1.908                Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x AX014220   ..

 Align seg 1/1  to: AX014220  from: 1  to: 656

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||::::::::::|||      |||:::   :::       :::|||
      98 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 147

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
                 ::::::         ||||||      |||   |||:::::::
     148 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 197

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||::::::::::|||||||||:::::
     198 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 247

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         ::::::|||::::::   :::||||||:::|||   |||::::::::::
     248 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 297

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         :::::::|||   |||      |||||||||   |||:::||||||||||
     298 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 346

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
                 :::::::::|||   :::               :::
     347 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 388

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      ||||||:::|||
     389 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 421

seq_name: gb_ro1:MCCOLIPAS

seq_documentation_block:
LOCUS       MCCOLIPAS     490 bp    mRNA            ROD       21-NOV-1996
DEFINITION  M.coypus mRNA for colipase.
ACCESSION   X82998
VERSION     X82998.1  GI:599866
KEYWORDS    colipase.
SOURCE      nutria.
  ORGANISM  Myocastor coypus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Hystricognathi; Myocastoridae;
            Myocastor.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Thirstrup,K., Carriere,F., Hjorth,S.A., Rasmussen,P.B.,
            Nielsen,P.F., Ladefoged,C., Thim,L. and Boel,E.
  TITLE     Cloning and expression in insect cells of two pancreatic lipases
            and a procolipase from Myocastor coypus
  JOURNAL   Eur. J. Biochem. 227 (1-2), 186-193 (1995)
  MEDLINE   95154288
REFERENCE   2  (bases 1 to 490)
  AUTHORS   Thirstrup,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1994) K. Thirstrup, Univ. Hospital, Lab. for
            Molecular Endocrinology, Rigshospitalet 6321, Blegdamsvej 9,
            DK-2100 Copenhagen, DENMARK
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Myocastor coypus"
                     /db_xref="taxon:10157"
                     /tissue_type="pancreas"
     CDS             46. .381
                     /codon_start=1
                     /product="colipase"
                     /protein_id="CAA58119.1"
                     /db_xref="GI:599867"
                     /db_xref="SWISS-PROT:P42889"
                     /translation="MEKVLALVLLTLAVAYAAPDPRGLIINLDNGELCLNSAQCKSQC
                     CQHDSPLGLARCADKARENSGCSPQTIYGIYYLCPCERGLTCDGDKSIIGAITNTNYG
                     ICQDPQSKK"
     sig_peptide     46. .96
     mat_peptide     97. .378
                     /note="putative propeptide: 97. .111"
                     /product="colipase"
     misc_feature    97. .111
                     /note="putative propeptide"
                     /product="colipase"
BASE COUNT       96 a    181 c    120 g     93 t
ORIGIN


alignment_scores:
            Quality:  134.50             Length:     95
              Ratio:  2.360                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 34.737

alignment_block:
 US-09-215-435-179 x MCCOLIPAS   ..

 Align seg 1/1  to: MCCOLIPAS  from: 1  to: 490

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         |||:::      |||   |||:::::::::   :::|||::::::   ||
     127 CTGGACAACGGCGAGCTCTGCCTGAACAGCGCCCAGTGCAAGAGTCAGTG 176

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         |||||||:::                  ::::::||||||:::|||:::
     177 CTGCCAGCACGACAGCCCCCTGGGCCTGGCCCGCTGCGCCGACAAGGCCA 226

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
           |||:::|||   |||:::   |||::::::::::|||   |||
     227 GAGAGAACAGCGGCTGCTCCCCACAGACGATCTATGGAATCTACTACTTG 276

      83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn......Gl 97
         |||||||||   |||:::|||||||||      :::|||:::
     277 TGTCCCTGCGAGCGGGGCCTGACCTGTGACGGGGACAAGTCCATCATCGG 326

      97 uLysTrpLeuSerIleAlaTyrGlyArgCysGln..............L 109
                 :::      ||||||   ||||||              |
     327 CGCCATCACCAACACCAACTACGGCATCTGCCAGGATCCTCAGTCCAAGA 376

     109 ysIleGlyArgGlnLysLeuAlaLysLysMetPhe 120
         ||   |||::::::   ||||||   :::   |||
     377 AGTGAGGCCACCACGCGCTGGCCCTGTCCCCTTTC 411

seq_name: gb_om:ECCOLIPB

seq_documentation_block:
LOCUS       ECCOLIPB      497 bp    mRNA            MAM       11-NOV-1994
DEFINITION  E.caballus mRNA for colipase B.
ACCESSION   X74344
VERSION     X74344.1  GI:572678
```

```
KEYWORDS    colipase.
SOURCE      horse.
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Kerfelec,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1993) B. Kerfelec, CNRS, CBM, 31 Chemin Joseph
            Aiguier, 13402 Marseille, Cedex 9, FRANCE
REFERENCE   2  (bases 1 to 497)
  AUTHORS   Crenon,I., Granon,S., Chapus,C. and Kerfelec,B.
  TITLE     Molecular cloning and expression of two horse pancreatic cDNA
            encoding colipase A and B
  JOURNAL   Biochim. Biophys. Acta 1213 (3), 357-360 (1994)
  MEDLINE   94325330
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /tissue_type="pancreas"
     CDS             <1. .327
                     /codon_start=1
                     /product="colipase B"
                     /protein_id="CAA52391.1"
                     /db_xref="GI:572679"
                     /db_xref="SWISS-PROT:P02705"
                     /translation="LALLLVALAVAYAVPDPRGVIINLEAGEICMNSAQCKSECCHRE
                     SSLSLARCAAKASENSECSAWTLYGVYYKCPCERGLTCQVDKTLVGSIMNTNFGICFD
                     AARSEE"
BASE COUNT      118 a    161 c    125 g     93 t
ORIGIN


alignment_scores:
            Quality:  134.00            Length:     63
              Ratio:  3.045               Gaps:      0
 Percent Similarity:  69.841   Percent Identity: 39.683

alignment_block:
 US-09-215-435-179 x ECCOLIPB   ..

 Align seg 1/1  to: ECCOLIPB  from: 1  to: 497

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         ||||||      |||   |||:::::::::   :::|||::::::   ||
      70 TTGGAGGCGGGCGAGATCTGCATGAACAGCGCCCAGTGCAAGAGCGAATG 119

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         ||||:::|||      :::         ::::::||||||   |||:::|
     120 CTGCCACCGGGAAAGCAGCCTGAGCCTGGCCCGCTGCGCAGCCAAAGCCA 169

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
         ||||||:::|||   |||:::::::   :::::::::|||   |||
     170 GCGAGAACAGCGAGTGCTCTGCCTGGACACTCTACGGGGTTTACTACAAG 219

      83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLys 95
         |||||||||   |||:::||||||||||     :::|||
     220 TGTCCCTGTGAGCGGGGCCTGACCTGCCAGGTGGACAAG 258

 seq_name: gb_om:AF148567

 seq_documentation_block:
 LOCUS       AF148567      510 bp    mRNA            MAM       06-MAY-2000
 DEFINITION  Sus scrofa pancreatic colipase mRNA, complete cds.
 ACCESSION   AF148567
 VERSION     AF148567.1  GI:7711135
 KEYWORDS    .
 SOURCE      pig.
   ORGANISM  Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE   1  (bases 1 to 510)
   AUTHORS   Darnis,S. and Chaix,J.C.
   TITLE     Cloning, sequencing and functional expression of porcine pancreatic
             colipase
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 510)
   AUTHORS   Darnis,S. and Chaix,J.C.
   TITLE     Direct Submission
   JOURNAL   Submitted (04-MAY-1999) LBBN, CNRS.ESA6033, Av Escadrille Normandie
             Niemen, Marseille 13397, France
 FEATURES             Location/Qualifiers
      source          1. .510
                      /organism="Sus scrofa"
                      /db_xref="taxon:9823"
                      /tissue_type="pancreas"
      CDS             44. .382
                      /note="cofactor; triglyceride lipase"
                      /codon_start=1
                      /product="colipase"
                      /protein_id="AAF67823.1"
                      /db_xref="GI:7711136"
                      /translation="MEKVLALLLVTLTVAYAVPDPRGIIINLDEGELCLNSAQCKSNC
                      CQHDTILSLSRCALKARENSECSAFTLYGVYYKCPCERGLTCEGDKSLVGSITNTNFG
                      ICHDVGRSSD"
 BASE COUNT      117 a    160 c    121 g    112 t
 ORIGIN


 alignment_scores:
             Quality:  133.00            Length:     85
               Ratio:  2.463               Gaps:      2
  Percent Similarity:  63.529   Percent Identity: 34.118

 alignment_block:
  US-09-215-435-179 x AF148567   ..

  Align seg 1/1  to: AF148567  from: 1  to: 510

       33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
          |||:::      |||   |||:::::::::   :::|||:::::::::||
      125 CTGGATGAGGGCGAGCTCTGCCTGAACAGTGCCCAGTGCAAGAGCAACTG 174

       49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
          ||||||||:::                 |||:::|||||   |||:::
      175 CTGCCAGCATGACACAATCCTGAGCCTGTCCCGCTGCGCACTCAAGGCCA 224

       66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
          |||:::|||   |||:::::::   :::::::::|||   |||
      225 GAGAGAACAGCGAGTGTTCTGCCTTCACGCTCTATGGGGTTTACTACAAG 274

       83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 99
          |||||||||   |||:::|||||||||         ::::::::|||
      275 TGTCCCTGTGAACGGGGCCTGACCTGT.........GAGGGGACAAGAG 315

       99 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 111
           |||                :::      :::|||   |||:::   :::|
      316 TCTCGTGGGCTCCATCACCAACACCAACTTTGGTATCTGCCATGATGTTG 365

      111 lyArg 112
          |||||
      366 GACGC 370

  seq_name: gb_om:ECCO1IPA

  seq_documentation_block:
  LOCUS       ECCO1IPA      472 bp    mRNA            MAM       28-JUL-1995
  DEFINITION  E.caballus mRNA for colipase A.
  ACCESSION   X74503
  VERSION     X74503.1  GI:572676
  KEYWORDS    anchoring protein; colipase A; secreted protein; triacylglycerol
              digestion.
  SOURCE      horse.
    ORGANISM  Equus caballus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Kerfelec,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1993) B. Kerfelec, CNRS, CBM, 31 Chemin Joseph
            Aiguier, 13402 Marseille, Cedex 9, FRANCE
REFERENCE   2  (bases 1 to 472)
  AUTHORS   Crenon,I., Granon,S., Chapus,C. and Kerfelec,B.
  TITLE     Molecular cloning and expression of two horse pancreatic cDNA
            encoding colipase A and B
  JOURNAL   Biochim. Biophys. Acta 1213 (3), 357-360 (1994)
  MEDLINE   94325330
COMMENT     conflicts with previous reported amino acid sequence (Pierrot et
            al., 1982, Eur.J.Biochem. 123, 347-354:E22->Q22; E32->S32;
            N89->D89; K92->R92; S/E93->S93; R94->E94; E95.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /clone_lib="cDNA library in lambda gt10"
                     /tissue_type="pancreas"
     sig_peptide     <1. .33
     CDS             <1. .321
                     /codon_start=1
                     /product="colipase A precursor"
                     /protein_id="CAA52611.1"
                     /db_xref="GI:572677"
                     /db_xref="SWISS-PROT:P02704"
                     /translation="LLLVALAVAYAVPDPRGVIINLEAGEICLNSAQCKSECCHQESS
                     LSLARCAAKASENSECSAWTLYGVYYKCPCERGLTCQVDKTLVGSIMNTNFGICFDAA
                     RSEE"
     mat_peptide     34. .318
                     /gene="colipase A"
                     /product="colipase A"
     gene            34. .318
                     /gene="colipase A"
     misc_feature    196. .210
                     /gene="colipase A"
                     /note="interfacial recognition site"
     3'UTR           319. .471
     polyA_signal    441. .447
BASE COUNT       99 a    158 c    123 g     92 t
ORIGIN


alignment_scores:
            Quality:  131.00            Length:     63
              Ratio:  2.977               Gaps:      0
 Percent Similarity:  69.841   Percent Identity: 38.095

alignment_block:
 US-09-215-435-179 x ECCO1IPA   ..

 Align seg 1/1  to: ECCO1IPA  from: 1  to: 472

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         ||||||      |||   |||:::::::::   :::|||::::::   ||
      64 TTGGAGGCGGGCGAGATCTGCCTGAACAGCGCCCAGTGCAAGAGCGAATG 113

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         ||||::::::      :::         ::::::||||||   |||:::|
     114 CTGCCACCAGGAAAGCAGCCTGAGCCTGGCCCGCTGTGCAGCCAAAGCCA 163

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
         |||||:::|||   |||::::::   :::::::::|||   |||
     164 GCGAGAACAGCGAGTGCTCTGCCTGGACACTCTACGGGGTTTACTACAAG 213

      83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLys 95
         |||||||||   |||:::|||||||||      :::|||
     214 TGTCCCTGTGAGCGGGGCCTGACCTGCCAGGTGGACAAG 252

seq_name: gb_ro2:RATCOLQ
```

```
seq_documentation_block:
LOCUS       RATCOLQ       492 bp    mRNA            ROD       04-AUG-1994
DEFINITION  Rat colipase mRNA, complete cds.
ACCESSION   M58370
VERSION     M58370.1  GI:203504
KEYWORDS    colipase.
SOURCE      Rat, cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Payne,R.M., Sims,H.F., Jennens,M.L. and Lowe,M.E.
  TITLE     Rat pancreatic lipase and two related proteins: enzymatic
            properties and mRNA expression during development
  JOURNAL   Am. J. Physiol. 266 (5 Pt 1), G914-G921 (1994)
  MEDLINE   94262798
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             59. .397
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA20505.1"
                     /db_xref="GI:203505"
                     /translation="MKVLVVLLVTLVAVAYAAPGPRGLFINLEDGEICVNSMQCKSRC
                     CQHDTILGIARCTHKAMENSECSPKTLYGIYYRCPCERGLTCEGDRSIIGAITNTNYG
                     VCLDSTRSKQ"
BASE COUNT      107 a    163 c    119 g    103 t
ORIGIN


alignment_scores:
            Quality:  122.50            Length:    122
              Ratio:  1.828               Gaps:      3
 Percent Similarity:  54.918   Percent Identity: 27.869

alignment_block:
 US-09-215-435-179 x RATCOLQ   ..

 Align seg 1/1  to: RATCOLQ  from: 1  to: 492

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         :::|||      :::   :::|||   :::|||:::      |||
      47 GTACCAGTCACCATGAAGGTCCTTGTTGTTCTGCTTGTAACCCTCGTTGC 96

      19 uLeuThrArgGlySerLeuSerProThrLysTyrAsnLeu...LeuGluL 35
         ::::::   ::::::   :::|||           :::   ||||||
      97 CGTGGCCTATGCAGCTCCTGGACCCCGGGGTCTTTTTATCAACCTGGAGG 146

      35 euLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGln 51
            |||   |||:::::::::::::::|||::::::   |||||||||
     147 ACGGTGAGATCTGCGTAAACAGTATGCAGTGTAAGAGCAGATGCTGCCAA 196

      52 ArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluGl 68
         :::                 ::::::|||:::::::|||:::   |||::
     197 CATGACACCATCCTGGGCATCGCCCGGTGCACACACAAGGCCATGGAGAA 246

      68 ySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProC 85
         :|||   |||:::   :::::::::::::|||   |||       |||||||
     247 CAGCGAGTGCTCCCCAAAGACCCTCTATGGGATCTACTACAGGTGTCCCT 296

      85 ysLeuArgAsnLeuThrCys.........................Ile 92
         ||   |||:::|||||||||                          |||
     297 GTGAGCGGGGCCTGACCTGTGAGGGGGACAGGAGCATCATTGGCGCCATC 346

      93 TyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCysGlnLy 109
         :::   |||                  ||||||   |||
     347 ACCAACACCAAC.....................TACGGCGTCTGCCTCGA 375

     109 sIleGlyArgGlnLys 114
             |||:::|||
```

```
    376 CTCCACCCGCTCCAAG 391

seq_name: gb_ro2:RATCOLIP

seq_documentation_block:
LOCUS       RATCOLIP      446 bp    mRNA            ROD       27-APR-1993
DEFINITION  Rat colipase mRNA, complete cds.
ACCESSION   M33333
VERSION     M33333.1  GI:203502
KEYWORDS    colipase.
SOURCE      Rattus norvegicus (strain Wistar) (library: puc) pancreas cDNA to
            mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Wicker,C. and Puigserver,A.
  TITLE     Rat pancreatic colipase mRNA: Nucleotide sequence of a cDNA clone
            and nutritional regulation by a lipidic diet
  JOURNAL   Biochem. Biophys. Res. Commun. 167, 130-136 (1990)
  MEDLINE   90179738
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /tissue_type="pancreas"
                     /tissue_lib="puc"
     mRNA            <1. .446
                     /note="colipase mRNA"
     CDS             18. .356
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA40943.1"
                     /db_xref="GI:203503"
                     /translation="MKVLVVLLVTLVAVAYAVPGPRGLFINLEDGEICVNSMQCKSRC
                     CQHDTILGIARCTHKAMENSECSPKTLYGIYYRCPCERGLTCEGDRSIIGAITNTNYG
                     VCLDSTRSKQ"
     sig_peptide     18. .68
     mat_peptide     69. .353
                     /product="colipase"
     polyA_signal    424. .430
     polyA_site      446
BASE COUNT       96 a    141 c    113 g     96 t
ORIGIN


alignment_scores:
            Quality:  122.00            Length:     91
              Ratio:  2.392               Gaps:      2
 Percent Similarity:  56.044   Percent Identity: 31.868

alignment_block:
 US-09-215-435-179 x RATCOLIP   ..

 Align seg 1/1  to: RATCOLIP  from: 1  to: 446

      33 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 49
         ||||||      |||   |||:::::::::::::::::|||:::::::   ||
      99 CTGGAGGACGGTGAGATCTGCGTAAACAGTATGCAGTGTAAGAGCAGATG 148

      49 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 66
         |||||||:::              :::::::|||:::::::|||:::
     149 CTGCCAACATGACACCATCCTGGGCATCGCCCGGTGCACACACAAGGCCA 198

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
           |||:::|||   |||:::   ::::::::::::|||   |||
     199 TGGAGAACAGCGAGTGCTCCCCAAAGACCCTCTATGGGATCTACTACAGG 248

      83 CysProCysLeuArgAsnLeuThrCys..................... 91
         |||||||||   |||:::||||||||||
     249 TGTCCCTGTGAGCGGGGCCTGACCTGTGAGGGGGACAGGAGCATCATTGG 298

      92 ....IleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 107
             |||   :::   |||                 ||||||    |
     299 CGCCATCACCAACACCAAC.....................TACGGCGTCT 327

     107 ysGlnLysIleGlyArgGlnLys 114
         ||            |||:::|||
     328 GCCTCGACTCCACCCGCTCCAAG 350

seq_name: gb_in3:CEW02A2

seq_documentation_block:
LOCUS       CEW02A2     35879 bp    DNA             INV       27-OCT-2000
DEFINITION  Caenorhabditis elegans cosmid W02A2, complete sequence.
ACCESSION   Z82286
VERSION     Z82286.1  GI:1914563
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 35879)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
            Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 35879)
  AUTHORS   Ainscough,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Mar 30, 1997 this sequence version replaced gi:1666624.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            This sequence is the entire insert of clone W02A2. The start of
            this sequence (1. .106) overlaps with the end of sequence AL132901.
            The end of this sequence (35780. .35879) overlaps with the start of
            sequence AL132865.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=W02A2.
FEATURES             Location/Qualifiers
     source          1. .35879
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="W02A2"
     gene            join(830. .1480,2149. .2274,2875. .3228)
                     /gene="W02A2.1"
                     /note="fat-2"
     CDS             join(830. .1480,2149. .2274,2875. .3228)
                     /gene="W02A2.1"
                     /note="predicted using Genefinder
                     contains similarity to Pfam domain: PF00487 (Fatty acid
                     desaturase), Score=108.7, E-value=5.7e-32, N=1
                     cDNA EST EMBL:Z14917 comes from this gene
                     cDNA EST EMBL:M75805 comes from this gene
                     cDNA EST EMBL:M89208 comes from this gene
                     cDNA EST EMBL:M89244 comes from this gene
                     cDNA EST yk13h1.5 comes from this gene
```

```
                cDNA EST yk13h1.3 comes from this gene
                cDNA EST yk173b1.5 comes from this gene
                cDNA EST yk271h6.3 comes from this gene
                cDNA EST yk276a2.3 comes from this gene
                cDNA EST yk437d4.3 comes from this gene
                cDNA EST yk276a2.5 comes from this gene
                cDNA EST yk271h6.5 comes from this gene
                cDNA EST yk437d4.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB05304.1"
                /db_xref="GI:4008402"
                /db_xref="SPTREMBL:Q9XUB8"
                /translation="MTIATKVNTNKKDLDTIKVPELPSVAAVKAAIPEHCFVKDPLTS
                ISYLIKDYVLLAGLYFAVPYIEHYLGWIGLLGWYWAMGIVGSALFCVGHDCGHGSFSD
                YEWLNDLCGHLAHAPILAPFWPWQKSHRQHHQYTSHVEKDKGHPWVTEEDYNNRTAIE
                KYFAVIPISGWLRWNPIYTIVGLPDGSHFWPWSRLFETTEDRVKCAVSGVACAICAYI
                AFVLCDYSVYTFVKYYYIPLLFQGLILVIITYLQHQNEDIEVYEADEWGFVRGQTQTI
                DRHWGFGLDNIMHNITNGHVAHHFFFTKIPHYHLLEATPAIKKALEPLKDTQYGYKRE
                VNYNWFFKYLHYNVTLDYLTHKAKGVLQYRSGVEAAKAKKAQ"
gene            join(5311. .5821,5870. .5913)
                /gene="W02A2.2"
CDS             join(5311. .5821,5870. .5913)
                /gene="W02A2.2"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAB05305.1"
                /db_xref="GI:4008403"
                /db_xref="SPTREMBL:Q9XUB7"
                /translation="MIRIFLVIALASVAVFSAPISRLPQSLDDIPAEFKELIPAKVTE
                YLKSITTEEKAAIKEFIKSVMGGNKSVEELSADIKERSPSLYAKVEKLDVLLRTKLAK
                LDPAALALFGKVIAQGLSFRQQFHAGYQPTPEMVKKLFKGYIAEYKTLSENAKATITD
                EFPIVVEFFQHEKIQAIIQQIVNY"
gene            complement(join(6733. .6777,6934. .7307,7357. .7828,
                7881. .7983,8042. .8198,8281. .8323))
                /gene="W02A2.3"
CDS             complement(join(6733. .6777,6934. .7307,7357. .7828,
                7881. .7983,8042. .8198,8281. .8323))
                /gene="W02A2.3"
                /note="predicted using Genefinder
                cDNA EST yk506f12.3 comes from this gene
                cDNA EST yk506f12.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB05306.1"
                /db_xref="GI:4008404"
                /db_xref="SPTREMBL:Q9XUB6"
                /translation="MARFLLIIGCATLAALASCESKSGEETPQYWRAILGDICQLPSF
                PRATGDPTRYVECVRQSSFAADRKDLGIWLLRECLPGYEFVASARRCKTVRSVNKQQE
                LCESANATEYNFCPTASAMQFLVEETREAPRQCACPNGEQNCVCPSPEILEPISAVKK
                VRRSAQQQPLRFSQYPSCPCPQNQPACTCSTNNQQNEQILVSISCCQQQETTTQAPVQ
                MQPCQCTPTQIQQNCQTTTTQQQYCPQQSAPVVQPQPCPLVQGSGVQNAQYQGICSWM
                IDPLATDPQSRTHYLQCQPAPNNLFCGRWQRMPCAPGTVFDVQSQVCVWDTNSQPGTL
                PTPAPYVSTQAPQNAQCGCTGGVQIGSCNQNYQCPGQSVCQVGQNAGQQCMVCCYFRK
                KARRF"
gene            join(17533. .17703,17782. .18180,18255. .18809,18906. .19283)
                /gene="W02A2.4"
CDS             join(17533. .17703,17782. .18180,18255. .18809,18906. .19283)
                /gene="W02A2.4"
                /note="predicted using Genefinder
                contains similarity to Pfam domain: PF00069 (Eukaryotic
                protein kinase domain), Score=194.9, E-value=4.1e-55, N=1"
                /codon_start=1
                /protein_id="CAB05307.1"
                /db_xref="GI:4008405"
                /db_xref="SPTREMBL:Q9XUB5"
                /translation="MMFGQTVEKLLKWDHAYLVRRAMQRNSKKVLCLSVKIDGKICHL
                TLNSNKDGWSCAKLLAKFPKITPKRFQHIHHLLDAWSLVVNSIVPVPRRQLVLQHNCI
                HLENILGNGAFGEVLKARYIPRGHTNGVEVAVKRALGTAQRDQMQEFCHEAQIMGTLK
                HDNVVKLYGIASLEQPIMLVMELITGGDLKKYLQSTPSIPNRQLIAFALDIANGMRHL
                VIKKVIHRDLAARNCLISRPTLQVKISDFGLAVQATEVIVKKLGKAPTRWLAPETFQK
                GIFNEKTDVWSFGVVLTEILTRCAADPLAPRTLEECKKYIIESPHPHRIENKEPKELA
                DLVEMCCDKNTICRANFQNAKRRLKIVIDQYDHPDGHKLELKKSEERKTLTIRRSTAT
                VTRKKSRDGGRQASVERKGHGTQRKVVSVLRKDRGALKLPSGIATNLEKPSPPPHTAK
                SYTIKSPTTPPVSQPPVSQPPVSQPPVSQPPVPQHPMTPPPASSARPHPK"
gene            join(19808. .19933,19980. .20072,20539. .20664,20720. .20822,
                20878. .20966,21275. .21528,21571. .21646,22430. .22684,
                22726. .22875,22948. .23135,24615. .24690,25406. .25525)
                /gene="W02A2.5"
CDS             join(19808. .19933,19980. .20072,20539. .20664,20720. .20822,
                20878. .20966,21275. .21528,21571. .21646,22430. .22684,
                22726. .22875,22948. .23135,24615. .24690,25406. .25525)
                /gene="W02A2.5"
                /note="predicted using Genefinder
                contains similarity to Pfam domain: PF01679
                (Uncharacterized protein family), Score=79.5,
                E-value=2.2e-20, N=1"
                /codon_start=1
                /protein_id="CAB05308.1"
                /db_xref="GI:4008406"
                /db_xref="SPTREMBL:Q9XUB4"
                /translation="MSTRRQLRVVLARNPPDAFANCPLFPTLDVILQCPFPGWTMEVL
                KMITDSLKWEIIPVVTPTIVGGLDWGTLQKNGTWSGVLGYLQNGTADAVALMYQKTDT
                RNEYFEYSYPVTNVQPVFVARKKTETLGSVLWNAFKPFSVEVWLCLLASLVLNLLIMI
                AISGIEFKLLFRSRFRPWEMLWHVVQLQLDEKSDDMLFYTLSGNIVLFVFALLQSGIL
                IEVYKGMLLTALITSNGDNPFANADEMIKLIGEKKYHLTTNYMGNWYFDDLQHSDQQH
                FVSLRAATASNPVIPAASVSAALDLVDSGKYIYPIQQDSLAMQMSKERCNYVYVSDGM
                PQVSSFFVFTKNFSGIAEFNTQIIMNQAFIQRTFNKYFNEGFKLGFIPKCEVTEPTAS
                DATKPLDIESVIGVFTIGALGVAASLMVFVFEIYHYWHVRILARRARMRDPWNVRNLA
                RIAQIHFTTHEQYSDIDVMRLVDAFNATSSSSDTSTFSTNQQTNMPITCTDIPKFICA
                LLLPPIGVWMEKGCGADLVINIVLTILGFIPGVIHACFIICWY"
gene            complement(join(26191. .26279,26334. .26431,26904. .27051,
                27549. .28993,29436. .29551,29600. .29796,29842. .30041,
                30091. .30143))
                /gene="W02A2.6"
CDS             complement(join(26191. .26279,26334. .26431,26904. .27051,
                27549. .28993,29436. .29551,29600. .29796,29842. .30041,
                30091. .30143))
                /gene="W02A2.6"
                /note="predicted using Genefinder
                cDNA EST yk35d2.3 comes from this gene
                cDNA EST yk35d2.5 comes from this gene
                cDNA EST yk178e5.5 comes from this gene
                cDNA EST yk182e9.5 comes from this gene
                cDNA EST yk182e9.3 comes from this gene"
                /codon_start=1
                /protein_id="CAB05309.1"
                /db_xref="GI:4008407"
                /db_xref="SPTREMBL:Q9XUB3"
                /translation="MVVSAEVIRKDAVFHVAWILGTGDSKKLSRREILDQNLPELCHS
                IIEMVPERHRGSATKTGLYLLSLLTYGTVLIHQVQVDFLKRDVEKLKELMKKKSFILL
                MAERFDRNQELQRKEDKFARLRSKPIMCVEELDRVDLAHLQAIGDELGINGNPGDFIM
                MDALPNMNQWIDNNSELNAIYGCVEPYLREKEITMHSTFVEGNGSNEHNKERRNDAVI
                ADFSQLLFPEIPEITLGEKFPIDVDSRKRSAILQEEQEEALQLPKEASEIVQEEPTKF
                VSIALLPSETVEQPAPQEPIQEPIQPIIEEEPAPQLELPQPELPPQLDAIDLVTIPASQ
                QDMVVEYLQLINDLPDDENSRLPPLPKDLELFEDVILPPPAKKSKVEEEEDALERARR
                RPSSRPVTPINQTDLTDLHSTVRPEDPSFAIDSQIHDVLPQRKKSKRNLPIIHSDDLE
                IDEAVQKVLQADYSSLVRKKEDVIAKIPPKTDAVAVLMNLPEPVFSIGYRLPPEVRDM
                FKACYNQAVGSPVSDDEEDEDEEEEEEYKYAKVCLLSPNRIVEDTLLLEEQPRQPEEF
                PSTDNINPPRQLQENPVFENLEYEAPPHPIRTARTPTPIKDLKYSVISLFPTPEKRRE
                TSIIAELNLDPIPVEEIDPLLTMRTEEELENVRRRQKSSLGVQFMRTDDLEEDTRRNR
                LFEDEERTRDAREDELFFYSSGSLLPNNRLNIHKELLNEAEARYPEWVNFNEFTADHD
                RKKAATAFEGLLLSLKNMKVEAKQEDPYFPILVRHISHEEM"
gene            join(30680. .30793,30842. .31362,31408. .31831,31882. .32030,
                32295. .32493)
                /gene="W02A2.7"
CDS             join(30680. .30793,30842. .31362,31408. .31831,31882. .32030,
                32295. .32493)
                /gene="W02A2.7"
                /note="predicted using Genefinder
                contains similarity to Pfam domain: PF00642 (Zinc finger

alignment_scores:
            Quality:   96.50              Length:    123
              Ratio:  1.485                Gaps:      6
 Percent Similarity:  52.846   Percent Identity: 31.707

alignment_block:
```

```
US-09-215-435-179 x CEW02A2/rev   ..

Align seg 1/1  to reverse of: CEW02A2  from: 1  to: 35879

      6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
        :::|||   |||:::   |||||||||:::|||||||||:::   |||
   7871 TACCTTTCACTCTTCAACCTTTTGTTTATCTTCTTGTTTATAGGAACTTT 7822

     22 g.GlySerLeuSerProThrLysTyrAsn..................... 31
          |||   :::   |||:::|||||||
   7821 GCGAGTCTGCCAATGCCACCGAGTACAACTTCTGCCCAACTGCCTCGGCC 7772

     32 .........LeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCy 45
                |||:::|||      :::::::::   |||:::      ||
   7771 ATGCAATTCCTCGTTGAGGAGACCCGTGAGGCTCCAAGACAGTGCGCCTG 7722

     45 sGluThrGly.........CysCysGlnArgAlaProAspAsnCysGluS 59
        |   :::|||         |||          :::|||:::      |||
   7721 TCCAAATGGAGAGCAAAACTGTGTCTGCCCATCGCCGGAAATCCTCGAGC 7672

     59 erHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVal 75
              |||   |||       :::   |||       |||       :::
   7671 CAATCTCTGCTGTCAAGAAGGTCCGCAGATCTGCCCAGCAACAGCCATTG 7622

     76 PhePheGlyGlnTyrArgAlaCysProCysLeuArgAsn.........Le 89
          |||:::||||||   :::||||||||||   :::|||
   7621 AGATTCTCCCAATACCCATCATGCCCATGCCCACAGAACCAACCAGCCTG 7572

     89 uThrCysIleTyrSer......LysAsnGluLysTrpLeuSerIleAlaT 104
          ||||||       :::       :::|||||||:::   |||   ::::::
   7571 CACATGCTCTACCAACAATCAGCAGAATGAGCAGATCCTT...GTCTCCA 7525

    104 yrGlyArgCysGlnLys 109
          :::   ||||||:::
   7524 TCTCCTGCTGCCAACAA 7508

seq_name: gb_htg24:CEY62E10

seq_documentation_block:
LOCUS       CEY62E10   254217 bp    DNA             HTG       06-SEP-1999
DEFINITION  Caenorhabditis elegans chromosome IV clone Y62E10, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   AL031580
VERSION     AL031580.13  GI:5823679
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 254217)
  AUTHORS   Lloyd,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Sep 5, 1999 this sequence version replaced gi:5763822.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .254217
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="Y62E10"
BASE COUNT    77287 a  44462 c  42975 g  79051 t  10442 others
ORIGIN


alignment_scores:
           Quality:   96.50            Length:    123
             Ratio:  1.485              Gaps:      6
 Percent Similarity:  52.846   Percent Identity: 31.707

alignment_block:
 US-09-215-435-179 x CEY62E10/rev   ..

 Align seg 1/1  to reverse of: CEY62E10  from: 1  to: 254217

      6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
        :::|||   |||:::   |||||||||:::|||||||||:::   |||
   2063 TACCTTTCACTCTTCAACCTTTTGTTTATCTTCTTGTTTATAGGAACTTT 2014

     22 g.GlySerLeuSerProThrLysTyrAsn..................... 31
          |||   :::   |||:::|||||||
   2013 GCGAGTCTGCCAATGCCACCGAGTACAACTTCTGCCCAACTGCCTCGGCC 1964

     32 .........LeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCy 45
                |||:::|||      :::::::::   |||:::      ||
   1963 ATGCAATTCCTCGTTGAGGAGACCCGTGAGGCTCCAAGACAGTGCGCCTG 1914

     45 sGluThrGly.........CysCysGlnArgAlaProAspAsnCysGluS 59
        |   :::|||         |||          :::|||:::      |||
   1913 TCCAAATGGAGAGCAAAACTGTGTCTGCCCATCGCCGGAAATCCTCGAGC 1864

     59 erHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVal 75
              |||   |||       :::   |||       |||       :::
   1863 CAATCTCTGCTGTCAAGAAGGTCCGCAGATCTGCCCAGCAACAGCCATTG 1814

     76 PhePheGlyGlnTyrArgAlaCysProCysLeuArgAsn.........Le 89
          |||:::||||||   :::||||||||||   :::|||
   1813 AGATTCTCCCAATACCCATCATGCCCATGCCCACAGAACCAACCAGCCTG 1764

     89 uThrCysIleTyrSer......LysAsnGluLysTrpLeuSerIleAlaT 104
          ||||||       :::       :::|||||||:::   |||   ::::::
   1763 CACATGCTCTACCAACAATCAGCAGAATGAGCAGATCCTT...GTCTCCA 1717

    104 yrGlyArgCysGlnLys 109
          :::   ||||||:::
   1716 TCTCCTGCTGCCAACAA 1700
```

OM of: US-09-215-435-179 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:28 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614	+	666.00	1331.22	3.5e-66	542	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408	+	666.00	1328.81	4.8e-66	680	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465	+	619.00	1238.97	4.9e-61	421	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ41401	+	135.50	257.75	2.2e-06	656	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98928	+	135.50	257.59	2.2e-06	666	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35367	-	86.00	142.00	6.13	2898		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28470	-	86.00	138.62	9.47	3988		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28469	-	86.00	137.43	11.02	4460		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF74198	+	84.50	154.32	1.26	683	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55249	+	83.00	137.27	11.25	2558		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55248	+	83.00	135.96	13.31	2895		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55252	+	83.00	135.87	13.46	2918		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55247	+	83.00	135.53	14.06	3014		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55253	+	83.00	135.43	14.25	3043		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55257	+	83.00	135.26	14.56	3092		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55250	+	83.00	135.11	14.84	3136		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55251	+	83.00	134.99	15.06	3170		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55254	+	83.00	134.72	15.60	3253		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55256	+	83.00	134.72	15.60	3253		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55255	+	83.00	134.72	15.61	3254		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55244	+	83.00	134.71	15.62	3255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55258	+	83.00	134.71	15.62	3255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55259	+	83.00	134.71	15.62	3255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55260	+	83.00	134.71	15.62	3255		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08843	+	82.50	151.24	1.88	624	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV07909	+	82.50	149.99	2.20	702	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75134	+	82.50	149.99	2.20	702	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56830	+	82.50	149.03	2.49	768	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08846	+	82.50	148.91	2.53	777	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08841	+	82.50	148.87	2.54	780	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08842	+	82.50	148.87	2.54	780	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75139	+	82.50	148.75	2.58	789	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56828	+	82.50	147.58	3.00	881	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA37042	+	82.50	133.98	17.15	3170		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF54250	+	82.50	133.98	17.15	3170		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75138	+	82.50	132.35	21.13	3696		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA44777	+	82.00	157.25	0.8675	322	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX22115	-	82.00	137.95	10.31	1984		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF16253	+	82.00	134.48	16.09	2751		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX22554	+	82.00	127.81	37.83	5154		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ98069	+	81.50	146.83	3.30	781		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37622	-	79.50	143.64	4.97	721		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34914	+	79.00	145.60	3.86	545		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ08867	-	78.50	129.72	29.63	221		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAT94120	-	78.50	127.75	38.12	266		

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614

seq_documentation_block:
ID AAX97614 standard; DNA; 542 BP.
XX
AC AAX97614;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 78.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; AAY35930.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 212; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX
SQ Sequence 542 BP; 132 A; 146 C; 126 G; 138 T; 0 other;

alignment_scores:
Quality: 666.00 Length: 121
Ratio: 5.504 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-179 x AAX97614 ..

Align seg 1/1 to: AAX97614 from: 1 to: 542

```
     1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

    17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    66 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATATAACCTTTTGG 115

    34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   116 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 165

    51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   166 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 215

    67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   216 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTC 265

    84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   266 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 315

   101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   316 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 365

   117 sLysMetPhePhe 121
       |||||||||||||
   366 GAAAATGTTCTTC 378
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408

seq_documentation_block:

```
ID   AAA26408 standard; cDNA; 680 BP.
XX
AC   AAA26408;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 63 SEQ ID NO:73.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW   osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW   antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW   immune disease; inflammation; blood disorder; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006698-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-US17130.
XX
PR   30-JUL-1998;   98US-0094657.
PR   05-AUG-1998;   98US-0095486.
PR   06-AUG-1998;   98US-0095454.
PR   06-AUG-1998;   98US-0095455.
PR   12-AUG-1998;   98US-0096319.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Rosen CA,  Ruben SM,  Duan R,  Moore PA,  Shi Y;
PI   Lafleur D,  Wei Y,  Ni J,  Florence KA,  Young PE,  Brewer LA;
PI   Soppet DR,  Endress GA,  Ebner R,  Olsen HS,  Mucenski M;
XX
DR   WPI; 2000-195282/17.
DR   P-PSDB; AAY91513.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders -
XX
PS   Claim 1; Page 416; 634pp; English.
XX
CC   The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC   human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC   antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC   antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC   corresponding secreted proteins are useful for preventing, treating or
CC   ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC   pathological conditions can be diagnosed by determining the amount of the
CC   proteins in a sample or by determining the presence of mutations in the
CC   polynucleotides. Specific uses are described for each of the
CC   polynucleotides, based on which tissues they are most highly expressed
CC   in, and include developing products for the diagnosis or treatment of
CC   cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, inflammation,
CC   allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC   osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC   transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC   cardiovascular disorders, reproductive disorders, gastrointestinal
CC   disorders, respiratory disorders and metabolic disorders. The proteins
CC   or polynucleotides can also be used as food additives or preservatives.
CC   The proteins are also useful for identifying their binding partners.
CC   AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 680 BP; 163 A; 185 C; 167 G; 156 T; 9 other;
```

alignment_scores:

```
            Quality:  666.00              Length:    121
              Ratio:  5.504                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-215-435-179 x AAA26408 ..

Align seg 1/1 to: AAA26408 from: 1 to: 680

```
     1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

    17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   183 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 232

    34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   233 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 282

    51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   283 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 332

    67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   333 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAGTATAGAGCGTGTC 382

    84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   383 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 432

   101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   433 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 482
```

```
     117 sLysMetPhePhe 121
         ||||||||||||
     483 GAAAATGTTCTTC 495

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465

seq_documentation_block:
ID   AAX51465 standard; cDNA; 421 BP.
XX
AC   AAX51465;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:44.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12687.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 177; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 421 BP; 96 A; 106 C; 110 G; 102 T; 7 other;


alignment_scores:
            Quality:  619.00            Length:    118
              Ratio:  5.383               Gaps:      0
 Percent Similarity:  97.458   Percent Identity: 95.763

alignment_block:
 US-09-215-435-179 x AAX51465   ..

 Align seg 1/1  to: AAX51465  from: 1  to: 421

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      62 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 111

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     112 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 161

      34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
         ||||||||||:::          ||||||||||||||||||||||||||
     162 AGCTCAAGGAGKSKKGCATSGGGAACCAGGACTGCGAGACTGGCTGCTGC 211

      51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     212 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 261

      67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     262 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTC 311

      84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     312 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 361

     101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
         ||||||||||||||||||||||||||||||||||||||||||||||||::
     362 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAG 411

     117 sLys 118
         :|||
     412 RAAA 415

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ41401

seq_documentation_block:
ID   AAZ41401 standard; cDNA; 656 BP.
XX
AC   AAZ41401;
XX
DT   19-JAN-2000  (first entry)
XX
DE   Human normal pancreas tissue derived cDNA 16.
XX
KW   Human; pancreas; cancer; treatment; anticancer; cytostatic;
KW   gene therapy; EST; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   DE19818598-A1.
XX
PD   21-OCT-1999.
XX
PF   19-APR-1998;   98DE-1018598.
XX
PR   19-APR-1998;   98DE-1018598.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-592039/51.
XX
PT   New nucleic acid sequences expressed in normal pancreatic tissues, and
PT   derived polypeptides, for treatment of pancreatic cancer and
PT   identification of therapeutic agents  -
XX
PS   Claim 3; Page 56; 92pp; German.
XX
CC   This invention describes novel cDNA sequences (A) that are highly
CC   expressed in normal human pancreatic tissue and which have anticancer
```

```
CC   and cytostatic activity. (A) are used (i) for recombinant expression of
CC   polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to
CC   identify agents suitable for treatment of pancreatic cancer; (ii)
CC   directly for treating this form of cancer (including expression from
CC   gene therapy vectors) and (iii) for generation of specific antibodies.
CC   (A) are identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, so
CC   should reduce the number of failures associated with the fact that ESTs
CC   from different libraries may represent different parts of the same
CC   unknown gene, distorting the estimated frequency of occurrence in a
CC   particular tissue. AAZ41386-Z41423 represent human normal pancreatic
CC   tissue derived cDNA fragments which encode the protein fragments
CC   represented in AAY59893-Y59920.
XX
SQ   Sequence 656 BP; 165 A; 200 C; 162 G; 129 T; 0 other;


alignment_scores:
           Quality:  135.50              Length:    111
             Ratio:  1.908                 Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x AAZ41401   ..

 Align seg 1/1  to: AAZ41401  from: 1  to: 656

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      98 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 147

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
               ::::::          ||||||      |||   |||:::::::
     148 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 197

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::
     198 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 247

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         :::::::|||::::::   :::|||||||:::|||   |||:::::::::::
     248 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 297

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         :::::::::|||   |||      |||||||||   |||:::||||||||||
     298 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 346

      92 leTyrSerLysAsnGluLysTrpLeu.............SerIleAla 103
                :::::::::|||   :::                 :::
     347 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 388

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      ||||||:::|||
     389 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 421

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98928

seq_documentation_block:
ID   AAC98928 standard; cDNA; 666 BP.
XX
AC   AAC98928;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:156.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54163.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 614; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 666 BP; 151 A; 217 C; 157 G; 133 T; 8 other;


alignment_scores:
           Quality:  135.50              Length:    111
             Ratio:  1.908                 Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x AAC98928   ..

 Align seg 1/1  to: AAC98928  from: 1  to: 666

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      91 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 140

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
               ::::::          ||||||      |||   |||:::::::
     141 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 190

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::
     191 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 240
```

```
      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         :::::::|||:::::::   :::|||||||:::|||   |||::::::::::::
     241 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 290

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         ::::::::|||   |||      |||||||||   |||:::|||||||||
     291 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 339

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
                :::::::::|||   :::              :::
     340 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 381

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      ||||||:::||| 
     382 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 414
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35367

seq_documentation_block:

```
ID   AAV35367 standard; cDNA; 2898 BP.
XX
AC   AAV35367;
XX
DT   01-OCT-1998  (first entry)
XX
DE   Human semaphorin encoding cDNA.
XX
KW   Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
KW   neurological disease; atopic skin inflammation; autoimmune disease;
KW   pain; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             370..2697
FT                   /*tag=  a
FT                   /product= "semaphorin"
XX
PN   WO9822504-A1.
XX
PD   28-MAY-1998.
XX
PF   12-NOV-1997;   97WO-JP04111.
XX
PR   15-NOV-1996;   96JP-0321068.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI   Furuyama T,  Inagaki S;
XX
DR   WPI; 1998-312416/27.
DR   P-PSDB; AAW63748.
XX
PT   Gene encoding new semaphorin nerve growth inhibitor - useful in
PT   diagnosis, treatment and study of neurological diseases
XX
PS   Claim 2; Page 37-39; 49pp; Japanese.
XX
CC   The present sequence encodes human semaphorin, a nerve growth
CC   inhibitor. The semaphorin protein, and gene encoding the protein,
CC   and their derivatives, are used in the diagnosis, treatment and
CC   study of neurological disorders such as atopic skin inflammation,
CC   autoimmune diseases and pain.
XX
SQ   Sequence 2898 BP; 794 A; 695 C; 748 G; 660 T; 1 other;
```

alignment_scores:

```
           Quality:   86.00             Length:     87
             Ratio:  1.536               Gaps:      7
 Percent Similarity:  64.368   Percent Identity: 35.632
```

alignment_block:

US-09-215-435-179 x AAV35367/rev ..

Align seg 1/1 to reverse of: AAV35367 from: 1 to: 2898

```
      16 PheLeuPheLeuLeuThrArgGlySerLeuSerProThrLysTyr.AsnL 32
         ::::::::::|||   :::|||:::|||||||||         ||| ||||
    1103 TACATTTTGTTGTCATCTCGGTCTTCGTTATCAGGAATCATATATGAACC 1054

      32 eu......LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGlu 46
         ||      |||  |||:::  ::::::  |||:::  :::||||||
    1053 TACAAATTTTGGTTCTTTCAGGAGCCGCTCATCGTCATGCTCAGTGCGAA 1004

      47 ThrGly...........CysCysGlnArgAlaProAspAsnCys...Gl 58
            |||           ||||||  |||:::|||:::  |||   ::
    1003 TATGGCCTAACTTCCCCATGCTGCGGAAGATCGCCGAGTCTCTGCCCCAA 954

      58 uSerHisCys...AlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
         :|||||||||   :::::::::   :::::::::::::|||:::   ||||
     953 TAGTCACTGTAGAGTCCAGCAAACAGCTCATTCCCAACTAGCGTGGACAC 904

      74 ln...ValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu 89
         ||       :::||||||        ||||||   |||   ::::::
     903 AAAGGAGGAGTTGGGGTCAAAAGGACATCTGCCCCTTCCTCTCTCAGATC 854

      90 ThrCysIle 92
            |||:::
     853 ...TGTGTG 848
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28470

seq_documentation_block:

```
ID   AAZ28470 standard; DNA; 3988 BP.
XX
AC   AAZ28470;
XX
DT   05-JAN-2000  (first entry)
XX
DE   Mouse semaphorin H variant (Sema Hv) polynucleotide sequence.
XX
KW   Semaphorin H; Sema H; Sema Hv; collapsin; bone structure formation;
KW   metastasis; cancer; antibody; drug screen; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             609..2943
FT                   /*tag=  a
FT                   /product= Sema_Hv
FT                   /note= "Mouse semaphorin H variant"
XX
PN   WO9947671-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-IB00495.
XX
PR   13-MAR-1998;   98US-0077997.
XX
PA   (LUKA/) LUKANIDIN E M.
PA   (CHRI/) CHRISTENSEN C R L.
XX
PI   Lukanidin EM,  Christensen CRL;
XX
DR   WPI; 1999-590975/50.
DR   P-PSDB; AAY43091.
XX
PT   New polypeptides and polynucleotides, useful in diagnosis and treatment
PT   of metastatic cancer  -
XX
PS   Claim 1; Page 85-91; 95pp; English.
XX
CC   This is the mouse Semaphorin-H variant (Sema Hv) polynucleotide sequence
CC   (Sema Hv). The Semaphorin/Collapsin family of molecules are characterised
```

```
CC   by a unique and highly conserved motif, within a  500 amino acid
CC   semaphorin domain. Some semaphorins exhibit inhibitory or repulsive
CC   functions in a neuronal context, and functions in bone structure
CC   formation are also implicated. It is thought that Sema H plays a role in
CC   metastasis. The invention uses the mouse Sema H gene in the diagnosis of
CC   metastatic cancer. Semaphorin polynucleotides and polypeptides are used
CC   in the methods of the invention, the polypeptides are useful for
CC   determining the metastatic potential of cells, by detecting their
CC   expression in biological samples. Antibodies specific for Sema H, are
CC   also useful therapeutically in inhibiting Sema-H polypeptide activity and
CC   therefore metastasis, and for purifying the polypeptides. Metastasis may
CC   also be inhibited by inhibiting the biological activity of the
CC   polypeptide using e.g. a small molecule inhibitor or a Semaphorin-H
CC   ligand (or fragment). The polynucleotides can also be used to inhibit
CC   polypeptide expression in cells using known antisense technology e.g. to
CC   prevent metastasis of cancer cells. They can be used to detect and
CC   quantify Sema-H mRNA levels in cells. The polypeptides, fusion proteins,
CC   multimeric proteins, antibodies or antisense oligonucleotides can be
CC   included in pharmaceutical compositions. The polynucleotides can be used
CC   to isolate similar sequences from other species and to produce mammalian
CC   cell lines and tumours with known metastatic potential, useful in
CC   anti-metastatic drug screening.
XX
SQ   Sequence 3988 BP; 1136 A; 873 C; 955 G; 1024 T; 0 other;


alignment_scores:
            Quality:   86.00              Length:     87
              Ratio:  1.536                Gaps:      7
 Percent Similarity:  64.368   Percent Identity: 35.632

alignment_block:
 US-09-215-435-179 x AAZ28470/rev   ..

 Align seg 1/1  to reverse of: AAZ28470  from: 1  to: 3988

      16 PheLeuPheLeuLeuThrArgGlySerLeuSerProThrLysTyr.AsnL 32
         ::::::::::|||   :::|||:::|||||||||         ||| ||||
    1343 TACATTTTGTTGTCATCTCGGTCTTCGTTATCAGGAATCATATATGAACC 1294

      32 eu......LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGlu 46
         ||      |||   |||:::   ::::::   |||:::   :::||||||
    1293 TACAAATTTTGGTTCTTTCAGGAGCCGCTCATCGTCATGCTCAGTGCGAA 1244

      47 ThrGly............CysCysGlnArgAlaProAspAsnCys...Gl 58
         |||            ||||||   |||:::|||:::   |||   ::
    1243 TATGGCCTAACTTCCCCATGCTGCGGAAGATCGCCGAGTCTCTGCCCCAA 1194

      58 uSerHisCys...AlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
         :|||||||||   :::::::::   :::::::::::::|||:::   ||||
    1193 TAGTCACTGTAGAGTCCAGCAAACAGCTCATTCCCAACTAGCGTGGACAC 1144

      74 ln...ValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu 89
         ||  .      :::|||||||          ||||||   |||   ::::::
    1143 AAAGGAGGAGTTGGGGTCAAAAGGACATCTGCCCCTTCCTCTCTCAGATC 1094

      90 ThrCysIle 92
            |||:::
    1093 ...TGTGTG 1088

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28469

 seq_documentation_block:
 ID   AAZ28469 standard; DNA; 4460 BP.
 XX
 AC   AAZ28469;
 XX
 DT   05-JAN-2000  (first entry)
 XX
 DE   Mouse semaphorin H (Sema H) polynucleotide sequence.
 XX
 KW   Semaphorin H; Sema H; Sema Hv; collapsin; bone structure formation;
 KW   metastasis; cancer; antibody; drug screen; ss.
 XX
 OS   Mus sp.
 XX
 FH   Key             Location/Qualifiers
 FT   CDS             609..2937
 FT                   /*tag= a
 FT                   /product= M_Sema_H
 FT                   /transl_except=(Pos:2452..2454, aa:Tyr)
 XX
 PN   WO9947671-A2.
 XX
 PD   23-SEP-1999.
 XX
 PF   12-MAR-1999;   99WO-IB00495.
 XX
 PR   13-MAR-1998;  98US-0077997.
 XX
 PA   (LUKA/) LUKANIDIN E M.
 PA   (CHRI/) CHRISTENSEN C R L.
 XX
 PI   Lukanidin EM,  Christensen CRL;
 XX
 DR   WPI; 1999-590975/50.
 DR   P-PSDB; AAZ28469.
 XX
 PT   New polypeptides and polynucleotides, useful in diagnosis and treatment
 PT   of metastatic cancer
 XX
 PS   Claim 1; Page 78-85; 95pp; English.
 XX
 CC   This is the mouse Semaphorin-H polynucleotide sequence (Sema H). The
 CC   Semaphorin/Collapsin family of molecules are characterised by a unique
 CC   and highly conserved motif, within a  500 amino acid semaphorin domain.
 CC   Some semaphorins exhibit inhibitory or repulsive functions in a neuronal
 CC   context, and functions in bone structure formation are also implicated.
 CC   It is thought that Sema H plays a role in metastasis. The invention uses
 CC   the mouse Sema H gene in the diagnosis of metastatic cancer. Semaphorin
 CC   polynucleotides and polypeptides are used in the methods of the
 CC   invention, the polypeptides are useful for determining the metastatic
 CC   potential of cells, by detecting their expression in biological samples.
 CC   Antibodies specific for Sema H, are also useful therapeutically in
 CC   inhibiting Sema-H polypeptide activity and therefore metastasis, and for
 CC   purifying the polypeptides. Metastasis may also be inhibited by
 CC   inhibiting the biological activity of the polypeptide using e.g. a small
 CC   molecule inhibitor or a Semaphorin-H ligand (or fragment). The
 CC   polynucleotides can also be used to inhibit polypeptide expression in
 CC   cells using known antisense technology e.g. to prevent metastasis of
 CC   cancer cells. They can be used to detect and quantify Sema-H mRNA levels
 CC   in cells. The polypeptides, fusion proteins, multimeric proteins,
 CC   antibodies or antisense oligonucleotides can be included in
 CC   pharmaceutical compositions. The polynucleotides can be used to isolate
 CC   similar sequences from other species and to produce mammalian cell lines
 CC   and tumours with known metastatic potential, useful in anti-metastatic
 CC   drug screening.
 XX
 SQ   Sequence 4460 BP; 1288 A; 950 C; 1023 G; 1199 T; 0 other;


 alignment_scores:
             Quality:   86.00              Length:     87
               Ratio:  1.536                Gaps:      7
  Percent Similarity:  64.368   Percent Identity: 35.632

 alignment_block:
  US-09-215-435-179 x AAZ28469/rev   ..

  Align seg 1/1  to reverse of: AAZ28469  from: 1  to: 4460

       16 PheLeuPheLeuLeuThrArgGlySerLeuSerProThrLysTyr.AsnL 32
          ::::::::::|||   :::|||:::|||||||||         ||| ||||
     1343 TACATTTTGTTGTCATCTCGGTCTTCGTTATCAGGAATCATATATGAACC 1294

       32 eu......LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGlu 46
```

```
      ||      |||   |||:::   ::::::   |||:::   :::|||||| 
 1293 TACAAATTTTGGTTCTTTCAGGAGCCGCTCATCGTCATGCTCAGTGCGAA 1244

   47 ThrGly.....CysCysGlnArgAlaProAspAsnCys...Gl 58
      |||           ||||||   |||:::|||:::   |||   :: 
 1243 TATGGCCTAACTTCCCCATGCTGCGGAAGATCGCCGAGTCTCTGCCCCAA 1194

   58 uSerHisCys...AlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
      :|||||||||   :::::::::   ::::::::::::|||:::   ||||
 1193 TAGTCACTGTAGAGTCCAGCAAACAGCTCATTCCCAACTAGCGTGGACAC 1144

   74 ln...ValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeu 89
      ||        :::|||||||        ||||||   |||   ::::::
 1143 AAAGGAGGAGTTGGGGTCAAAAGGACATCTGCCCCTTCCTCTCTCAGATC 1094

   90 ThrCysIle 92
      |||:::
 1093 ...TGTGTG 1088

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF74198

seq_documentation_block:
ID   AAF74198 standard; DNA; 683 BP.
XX
AC   AAF74198;
XX
DT   02-MAY-2001  (first entry)
XX
DE   DNA encoding evironmental stress tolerant protein SEQ ID 23.
XX
KW   Environmental stress resistance; salt; heat; desert; transgenic plant;
KW   ds.
XX
OS   Salsola komarovii.
XX
PN   WO200106006-A1.
XX
PD   25-JAN-2001.
XX
PF   19-JUL-2000; 2000WO-JP04862.
XX
PR   19-JUL-1999;   99JP-0235910.
PR   24-MAR-2000; 2000JP-0085377.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Yamada A,  Ozeki Y,  Saito T;
XX
DR   WPI; 2001-147355/15.
DR   P-PSDB; AAB80619.
XX
PT   Screening method to obtain DNA encoding environmental stress resistance
PT   factor, useful for producing transgenic plants resistant to
PT   environmental stress -
XX
PS   Claim 47; Page 105-106; 167pp; Japanese.
XX
CC   Polynucleotide sequences AAF74187 - AAF74218 encode proteins
CC   AAB80608 - AAB80639, which impart environmental stress resistance. The
CC   invention relates to a method for identifying DNA encoding proteins
CC   imparting environmental stress resistance. The method comprises inserting
CC   cDNA from a library originating in a salt-resistant organism into a host
CC   cell, culturing the transformants under conditions in which the
CC   untransformed host does not grow well, and selecting for viable clones.
CC   The method is useful for obtaining DNA encoding environmental stress
CC   resistance factors. The DNA encoding proteins conferring environmental
CC   stress resistance, can be used in the production of plants resistant to
CC   environmental stress, which can be cultivated in unfavourable
CC   environments such as deserts, salt damaged ground, cold regions and the
CC   oceans. They can be used for increasing the area of land covered by green
CC   plants, and desert greening and afforestation, in order to counter the
CC   effects of the increase in atmospheric carbon dioxide concentration. PCR
CC   primers AAF74219 and AAF74220 are used in an example illustrating the
CC   method of the invention.
XX
SQ   Sequence 683 BP; 198 A; 133 C; 127 G; 225 T; 0 other;


alignment_scores:
            Quality:   84.50              Length:     99
              Ratio:  1.760                Gaps:      4
 Percent Similarity:  48.485   Percent Identity: 30.303

alignment_block:
 US-09-215-435-179 x AAF74198   ..

 Align seg 1/1  to: AAF74198  from: 1  to: 683

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         ::::::   :::   ||||||   ||||||:::|||   |||
      66 CTAATTGCTTCTCTACTTCTTTCTCTCTTTGTTCTTCAGTTTGTTCATGC 115

      18 eLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluL 35
           :::         :::|||   :::   :::       |||
     116 AGTTGAACCTATTTCATCCTCCAATCAAGTGGGTAGCAACACTGGAGGTA 165

      35 euLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGln 51
           :::|||||   :::         ||||||   :::::|||
     166 CCTCAGAGAGTAAAGTG.....GATTGTGGGGCGGCATGTACGGTG 206

      52 ArgAla.............ProAspAsnCysGluSerHisCysAlaGl 63
         |||:::              |||:::   |||:::      |||
     207 AGGTGCAGCGCCTCGAAGAGGCCAAACCTATGCAACAGGTCATGT..... 251

      63 uLysGlySerGluGlySerLeuCysGlnThr.............GlnV 75
                   ||||||   |||:::|||
     252 ............GGCAGTTGTTGCAAGACGTGCAACTGCGTGCCACCAG 288

      75 alPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThr 90
               |||:::|||:::||||||||||||      |||||||||
     289 GCACTTCCGGCAACTACGAAGCCTGCCCTTGTTACGCCAACTTGACC 335

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55249

seq_documentation_block:
ID   AAF55249 standard; cDNA; 2558 BP.
XX
AC   AAF55249;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of murine parkin2 with a frameshift mutation.
XX
KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
KW   brain tumour; head trauma; stroke; vascular irregularity;
KW   metabolic irregularity; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..320
FT                   /*tag= a
FT                   /product= "truncated parkin2"
XX
PN   EP1081225-A1.
XX
PD   07-MAR-2001.
XX
PF   30-AUG-1999;   99EP-0116766.
XX
PR   30-AUG-1999;   99EP-0116766.
XX
PA   (BIOF-) BIOFRONTERA PHARM GMBH.
XX
```

```
PI   Luebbert H;
XX
DR   WPI; 2001-212797/22.
DR   P-PSDB; AAB67522.
XX
PT   New polynucleotides encoding mouse parkin2 protein, useful for
PT   producing a transgenic non-human animal as an animal model for
PT   neurodegenerative diseases  -
XX
PS   Claim 3; Page 24-25; 62pp; English.
XX
CC   The present sequence encodes a murine parkin2 polypeptide. The
CC   polynucleotide sequence contains a frameshift mutation, leading to a
CC   truncated protein. Mutations or deletions in the parkin2 gene cause
CC   Parkinson's disease in humans. The human parkin2 gene is located in
CC   gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are
CC   useful for analysing neurodegenerative diseases. They are also useful
CC   for testing the efficacy of the treatment of a neurodegenerative disease
CC   such as Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC   Pick's disease, Prion disease, and secondary causes inducing
CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
CC   stroke, vascular irregularities or metabolic irregularities, associated
CC   with a less active or non-active parkin protein.
XX
SQ   Sequence 2558 BP; 674 A; 625 C; 630 G; 629 T; 0 other;


alignment_scores:
            Quality:   83.00            Length:    106
              Ratio:  1.566              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55249    ..

 Align seg 1/1  to: AAF55249  from: 1  to: 2558

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
         :::|||:::::::|||:::|||||||||||||||||:::|||||||:::::
    1876 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTGGTGGTGGT 1925

      22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
          |||       :::|||   |||          |||   :::|||:::
    1926 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 1975

      37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
         ||||||                                 |||::::::
    1976 TATCTTGTGG.........................TGTGAGGAAAGG 1997

      54 ProAspAsnCysGluSerHisCysAlaGluLysGly.............. 65
         |||   :::   :::|||||||   :::   ::::||
    1998 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2047

      66 ....................................SerGluGlySerL 70
                                           :::       ||||
    2048 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2097

      70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
         || |||:::   ::::::||||||   ::::::   ||||||   |||||
    2098 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2147

      86 uArgAsnLeuThrCys 91
         |          |||
    2148 CCTTCCACACCCGTGT 2163

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55248

 seq_documentation_block:
 ID   AAF55248 standard; cDNA; 2895 BP.
 XX
 AC   AAF55248;
 XX
 DT   29-MAY-2001  (first entry)
 XX
 DE   Nucleotide sequence of a murine truncated parkin2 polypeptide.
 XX
 KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
 KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
 KW   brain tumour; head trauma; stroke; vascular irregularity;
 KW   metabolic irregularity; ss.
 XX
 OS   Mus sp.
 XX
 FH   Key             Location/Qualifiers
 FT   CDS             129..1163
 FT                   /*tag=  a
 FT                   /product= "truncated parkin2"
 XX
 PN   EP1081225-A1.
 XX
 PD   07-MAR-2001.
 XX
 PF   30-AUG-1999;   99EP-0116766.
 XX
 PR   30-AUG-1999;   99EP-0116766.
 XX
 PA   (BIOF-) BIOFRONTERA PHARM GMBH.
 XX
 PI   Luebbert H;
 XX
 DR   WPI; 2001-212797/22.
 DR   P-PSDB; AAB67521.
 XX
 PT   New polynucleotides encoding mouse parkin2 protein, useful for
 PT   producing a transgenic non-human animal as an animal model for
 PT   neurodegenerative diseases  -
 XX
 PS   Claim 3; Page 23-24; 62pp; English.
 XX
 CC   The present sequence encodes a murine parkin2 polypeptide. The
 CC   polynucleotide sequence contains a deletion, leading to a truncated
 CC   protein. Mutations or deletions in the parkin2 gene cause Parkinson's
 CC   disease in humans. The human parkin2 gene is located in gene region
 CC   6q25.2-27. Parkin2 polypeptides and polynucleotides are useful for
 CC   analysing neurodegenerative diseases. They are also useful for testing
 CC   the efficacy of the treatment of a neurodegenerative disease such as
 CC   Parkinson's disease, Alzheimer's disease, Huntington's disease,
 CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
 CC   Pick's disease, Prion disease, and secondary causes inducing
 CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
 CC   stroke, vascular irregularities or metabolic irregularities, associated
 CC   with a less active or non-active parkin protein.
 XX
 SQ   Sequence 2895 BP; 749 A; 719 C; 712 G; 715 T; 0 other;


alignment_scores:
            Quality:   83.00            Length:    106
              Ratio:  1.566              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55248    ..

 Align seg 1/1  to: AAF55248  from: 1  to: 2895

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
         :::|||:::::::|||:::|||||||||||||||||:::|||||||:::::.
    2213 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTGGTGGTGGT 2262

      22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
          |||       :::|||   |||          |||   :::|||:::
    2263 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 2312
```

```
     37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
        ||||||                                 |||::::::
   2313 TATCTTGTGG.........................TGTGAGGAAAGG 2334

     54 ProAspAsnCysGluSerHisCysAlaGluLysGly.............. 65
        |||   :::   :::||||||   :::   :::|||
   2335 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2384

     66 .....................................SerGluGlySerL 70
                                            :::      ||||
   2385 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2434

     70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
        || |||:::   ::::::||||||   ::::::   ||||||   |||||
   2435 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2484

     86 uArgAsnLeuThrCys 91
        |            |||
   2485 CCTTCCACACCCGTGT 2500
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55252

seq_documentation_block:

```
ID   AAF55252 standard; cDNA; 2918 BP.
XX
AC   AAF55252;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of murine parkin2 with a frameshift mutation.
XX
KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
KW   brain tumour; head trauma; stroke; vascular irregularity;
KW   metabolic irregularity; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..680
FT                   /*tag=  a
FT                   /product= "truncated parkin2"
XX
PN   EP1081225-A1.
XX
PD   07-MAR-2001.
XX
PF   30-AUG-1999;   99EP-0116766.
XX
PR   30-AUG-1999;   99EP-0116766.
XX
PA   (BIOF-) BIOFRONTERA PHARM GMBH.
XX
PI   Luebbert H;
XX
DR   WPI; 2001-212797/22.
DR   P-PSDB; AAB67525.
XX
PT   New polynucleotides encoding mouse parkin2 protein, useful for
PT   producing a transgenic non-human animal as an animal model for
PT   neurodegenerative diseases  -
XX
PS   Claim 3; Page 28-29; 62pp; English.
XX
CC   The present sequence encodes a murine parkin2 polypeptide. The
CC   polynucleotide sequence contains a frameshift mutation, leading to a
CC   truncated protein. Mutations or deletions in the parkin2 gene cause
CC   Parkinson's disease in humans. The human parkin2 gene is located in
CC   gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are
CC   useful for analysing neurodegenerative diseases. They are also useful
CC   for testing the efficacy of the treatment of a neurodegenerative disease
CC   such as Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC   Pick's disease, Prion disease, and secondary causes inducing
CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
CC   stroke, vascular irregularities or metabolic irregularities, associated
CC   with a less active or non-active parkin protein.
XX
SQ   Sequence 2918 BP; 777 A; 724 C; 727 G; 690 T; 0 other;
```

```
alignment_scores:
            Quality:   83.00            Length:    106
              Ratio:  1.566              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55252   ..

 Align seg 1/1  to: AAF55252  from: 1  to: 2918

      6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
        :::|||::::::|||:::|||||||||||||||:::||||||::::::
   2236 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTTGGTGGTGGT 2285

     22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
         |||       :::|||   |||         |||   :::|||:::
   2286 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 2335

     37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
        ||||||                                 |||::::::
   2336 TATCTTGTGG.........................TGTGAGGAAAGG 2357

     54 ProAspAsnCysGluSerHisCysAlaGluLysGly.............. 65
        |||   :::   :::||||||   :::   :::|||
   2358 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2407

     66 .....................................SerGluGlySerL 70
                                            :::      ||||
   2408 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2457

     70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
        || |||:::   ::::::||||||   ::::::   ||||||   |||||
   2458 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2507

     86 uArgAsnLeuThrCys 91
        |            |||
   2508 CCTTCCACACCCGTGT 2523
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55247

seq_documentation_block:

```
ID   AAF55247 standard; cDNA; 3014 BP.
XX
AC   AAF55247;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of murine parkin2 with a frameshift mutation.
XX
KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
KW   brain tumour; head trauma; stroke; vascular irregularity;
KW   metabolic irregularity; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..446
FT                   /*tag=  a
FT                   /product= "truncated parkin2"
XX
PN   EP1081225-A1.
XX
```

```
PD   07-MAR-2001.
XX
PF   30-AUG-1999;   99EP-0116766.
XX
PR   30-AUG-1999;   99EP-0116766.
XX
PA   (BIOF-) BIOFRONTERA PHARM GMBH.
XX
PI   Luebbert H;
XX
DR   WPI; 2001-212797/22.
DR   P-PSDB; AAB67520.
XX
PT   New polynucleotides encoding mouse parkin2 protein, useful for
PT   producing a transgenic non-human animal as an animal model for
PT   neurodegenerative diseases  -
XX
PS   Claim 3; Page 22-23; 62pp; English.
XX
CC   The present sequence encodes a murine parkin2 polypeptide. The
CC   polynucleotide sequence contains a frameshift mutation in, leading to a
CC   truncated protein. Mutations or deletions in the parkin2 gene cause
CC   Parkinson's disease in humans. The human parkin2 gene is located in
CC   gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are
CC   useful for analysing neurodegenerative diseases. They are also useful
CC   for testing the efficacy of the treatment of a neurodegenerative disease
CC   such as Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC   Pick's disease, Prion disease, and secondary causes inducing
CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
CC   stroke, vascular irregularities or metabolic irregularities, associated
CC   with a less active or non-active parkin protein.
XX
SQ   Sequence 3014 BP; 781 A; 761 C; 734 G; 738 T; 0 other;


alignment_scores:
            Quality:   83.00              Length:    106
              Ratio:  1.566                Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55247   ..

 Align seg 1/1  to: AAF55247  from: 1  to: 3014

       6 TrpLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
         :::|||:::::::|||:::||||||||||||||||:::||||||:::::
    2332 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTTGGTGGTGGT 2381

      22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
          |||      :::|||   |||         |||   :::|||:::
    2382 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 2431

      37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
           ||||||                              |||::::::
    2432 TATCTTGTGG.........................TGTGAGGAAAGG 2453

      54 ProAspAsnCysGluSerHisCysAlaGluLysGly..............  65
         |||   :::   :::||||||   :::   :::|||
    2454 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2503

      66 .................................SerGluGlySerL 70
                                         :::      ||||
    2504 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2553

      70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
         || |||:::   ::::::||||||   ::::::   ||||||   |||||
    2554 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2603

      86 uArgAsnLeuThrCys 91
         |          |||
    2604 CCTTCCACACCCGTGT 2619
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55253

seq_documentation_block:
ID   AAF55253 standard; cDNA; 3043 BP.
XX
AC   AAF55253;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of murine parkin2 with a frameshift mutation.
XX
KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
KW   brain tumour; head trauma; stroke; vascular irregularity;
KW   metabolic irregularity; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..1019
FT                   /*tag= a
FT                   /product= "truncated parkin2"
XX
PN   EP1081225-A1.
XX
PD   07-MAR-2001.
XX
PF   30-AUG-1999;   99EP-0116766.
XX
PR   30-AUG-1999;   99EP-0116766.
XX
PA   (BIOF-) BIOFRONTERA PHARM GMBH.
XX
PI   Luebbert H;
XX
DR   WPI; 2001-212797/22.
DR   P-PSDB; AAB67526.
XX
PT   New polynucleotides encoding mouse parkin2 protein, useful for
PT   producing a transgenic non-human animal as an animal model for
PT   neurodegenerative diseases  -
XX
PS   Claim 3; Page 29-30; 62pp; English.
XX
CC   The present sequence encodes a murine parkin2 polypeptide. The
CC   polynucleotide sequence contains a frameshift mutation, leading to a
CC   truncated protein. Mutations or deletions in the parkin2 gene cause
CC   Parkinson's disease in humans. The human parkin2 gene is located in
CC   gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are
CC   useful for analysing neurodegenerative diseases. They are also useful
CC   for testing the efficacy of the treatment of a neurodegenerative disease
CC   such as Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC   Pick's disease, Prion disease, and secondary causes inducing
CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
CC   stroke, vascular irregularities or metabolic irregularities, associated
CC   with a less active or non-active parkin protein.
XX
SQ   Sequence 3043 BP; 806 A; 763 C; 737 G; 737 T; 0 other;


alignment_scores:
            Quality:   83.00              Length:    106
              Ratio:  1.566                Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55253   ..

 Align seg 1/1  to: AAF55253  from: 1  to: 3043
```

```
      6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
        :::|||:::::::|||:::|||||||||||||||||:::||||||:::::
   2361 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTGGTGGTGGT 2410

     22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
         |||      :::|||   |||         |||   :::|||:::
   2411 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 2460

     37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
          ||||||                        |||::::::
   2461 TATCTTGTGG.........................TGTGAGGAAAGG 2482

     54 ProAspAsnCysGluSerHisCysAlaGluLysGly.............. 65
        |||   :::   :::||||||   :::   :::|||
   2483 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2532

     66 ....................................SerGluGlySerL 70
                                            :::      ||||
   2533 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2582

     70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
        || |||:::   ::::::||||||   ::::::   ||||||   |||||
   2583 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2632

     86 uArgAsnLeuThrCys 91
        |            |||
   2633 CCTTCCACACCCGTGT 2648

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55257

seq_documentation_block:
ID   AAF55257 standard; cDNA; 3092 BP.
XX
AC   AAF55257;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of murine parkin2 with a frameshift mutation.
XX
KW   Parkin2; Parkinson's disease; 6q25.2-27; neurodegenerative disease;
KW   Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW   Multi-system atrophy; Wilson's disease; Pick's disease; Prion disease;
KW   brain tumour; head trauma; stroke; vascular irregularity;
KW   metabolic irregularity; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             129..173
FT                   /*tag= a
FT                   /product= "truncated parkin2"
XX
PN   EP1081225-A1.
XX
PD   07-MAR-2001.
XX
PF   30-AUG-1999;   99EP-0116766.
XX
PR   30-AUG-1999;   99EP-0116766.
XX
PA   (BIOF-) BIOFRONTERA PHARM GMBH.
XX
PI   Luebbert H;
XX
DR   WPI; 2001-212797/22.
DR   P-PSDB; AAB67530.
XX
PT   New polynucleotides encoding mouse parkin2 protein, useful for
PT   producing a transgenic non-human animal as an animal model for
PT   neurodegenerative diseases  -
XX
PS   Claim 3; Page 34-35; 62pp; English.
XX
CC   The present sequence encodes a murine parkin2 polypeptide. The
CC   polynucleotide sequence contains a frameshift mutation, leading to a
CC   truncated protein. Mutations or deletions in the parkin2 gene cause
CC   Parkinson's disease in humans. The human parkin2 gene is located in
CC   gene region 6q25.2-27. Parkin2 polypeptides and polynucleotides are
CC   useful for analysing neurodegenerative diseases. They are also useful
CC   for testing the efficacy of the treatment of a neurodegenerative disease
CC   such as Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, Multi-system atrophy, Wilson's disease,
CC   Pick's disease, Prion disease, and secondary causes inducing
CC   Parkinson's syndromes like toxins, drugs, brain tumours, head trauma,
CC   stroke, vascular irregularities or metabolic irregularities, associated
CC   with a less active or non-active parkin protein.
XX
SQ   Sequence 3092 BP; 818 A; 778 C; 761 G; 735 T; 0 other;


alignment_scores:
            Quality:   83.00              Length:    106
              Ratio:  1.566                Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 29.245

alignment_block:
 US-09-215-435-179 x AAF55257   ..

 Align seg 1/1  to: AAF55257  from: 1  to: 3092

      6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
        :::|||:::::::|||:::|||||||||||||||||:::||||||:::::
   2410 TTTTTATTTGTTTTGCTTTTGCTTTTTTTTTTTTTTTTTTTGGTGGTGGT 2459

     22 gGly......SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG 37
         |||      :::|||   |||         |||   :::|||:::
   2460 GGGGCTACCAAACTTGAGGCCTAGAGCTAAAAATCATATAGAAATGATGT 2509

     37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla 53
          ||||||                        |||::::::
   2510 TATCTTGTGG.........................TGTGAGGAAAGG 2531

     54 ProAspAsnCysGluSerHisCysAlaGluLysGly.............. 65
        |||   :::   :::||||||   :::   :::|||
   2532 CCAGCTGGCCTAAGTTCACACTTTTGTCCCAGTGGCCCTAGACTCCACCC 2581

     66 ....................................SerGluGlySerL 70
                                            :::      ||||
   2582 AGCCAGCTCCCAAAATGAAAAGACCACCTGTCAAGCAGCAGTCAGGAGTC 2631

     70 eu.CysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLe 86
        || |||:::   ::::::||||||   ::::::   ||||||   |||||
   2632 TGATGTCACCCATCACTATTTTTTTTCCATCATTGTGCTTGCCTCTGCCT 2681

     86 uArgAsnLeuThrCys 91
        |            |||
   2682 CCTTCCACACCCGTGT 2697
```

```
OM of: US-09-215-435-179 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  3:59 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2 +    86.50   69.69  3.1e+04   4403765
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-653-740-4 -    78.50  134.05  8.13      2663  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-09-073-594-4 -    78.50  134.05  8.13      2663  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-275-925-4 -    78.50  134.05  8.13      2663  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-085-199B-3 -    78.50  127.58  18.65     4796  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-142-078-63 -    77.50  154.04  0.6260     356  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-357-141-63 -    77.50  154.04  0.6260     356  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-11 +    76.00  126.57  21.21     3226  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-11 +    76.00  126.57  21.21     3226  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-11 +    76.00  126.57  21.21     3226
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-3 +    75.00  132.45  9.98      1556  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-3 +    75.00  132.45  9.98      1556  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-882-046-3 +    75.00  121.22  42.11     4315  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-181-706-1 +    75.00  120.27  47.60     4707  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-458-791-1 +    75.00  120.27  47.60     4707  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-459-066-1 +    75.00  120.27  47.60     4707  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-4 +    74.00  130.18  13.35     1573  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-4 +    74.00  130.18  13.35     1573  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-781-891-208 +    74.00  104.36  366.15    16442
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-085-199B-1 -    73.50  132.42  10.02     1164  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-860-091A-1 +    73.00  124.77  26.72     2116  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-9 +    73.00  117.69  66.29     4029  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-9 +    73.00  117.69  66.29     4029  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-9 +    73.00  117.69  66.29     4029
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-02766-15 +    72.00  141.66  3.06       375
/cgn1_7/ptodata/1/ina/backfiles1.seq:5428135-3 +    72.00  141.37  3.18       385  ! Paten
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-095-898-3 +    72.00  140.84  3.40       404  !  S
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-09439-3 +    72.00  140.84  3.40       404
/cgn1_7/ptodata/1/ina/backfiles1.seq:5194596-12 +    72.00  135.63  6.64       649  ! Pate
/cgn1_7/ptodata/1/ina/backfiles1.seq:5219739-12 +    72.00  135.63  6.64       649  ! Pate
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-02766-21 +    72.00  127.85  18.01     1316
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-197-636-1 +    71.50  112.53  128.39    4803  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-197-636-3 +    71.50  112.53  128.39    4803  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-619-554-1 +    71.50  107.40  247.83    7655  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-000-630C-1 +    71.00  122.82  34.31     1710  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-862-730C-1 +    71.00  122.82  34.31     1710  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-728-323A-1 -    71.00  114.97  93.85     3489  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-770-379-20 +    71.00   90.52  2.2e+03   32207  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-757-669A-20 +    71.00   90.52  2.2e+03   32207
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-543-238-1 +    70.50  136.76  5.74      437  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-420-526-1 +    70.50  136.76  5.74      437  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-551-437-4 +    70.50  121.59  40.16     1734
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-004-225-4 +    70.50  121.59  40.16     1734
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-084-346-4 +    70.50  121.59  40.16     1734
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-104-704-4 +    70.50  121.59  40.16     1734

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
           Quality:   86.50              Length:     73
             Ratio:  2.403                Gaps:      5
 Percent Similarity:  49.315   Percent Identity: 30.137

alignment_block:
 US-09-215-435-179 x US-09-103-840A-2   ..

 Align seg 1/1  to: US-09-103-840A-2  from: 1  to: 4403765

      39 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 55
         |||   |||||:::   |||:::      ||||||   |||:::|||
  3963587TGTTGCCGAAACCGATGTTGCCAATGCCGTTGTTGC...CGAAGCCCC.. 3963631

      55 pAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysG 72
         :::|||      :::|||:::::::::
  3963632.AGTTGCTGCTGCCGGTGTTCCCGGCGC..................... 3963658

      72 lnThrGlnValPhePheGlyGlnTyrArgAlaCysProCys..... 85
                                  |||:::|||   |||
  3963659.........................CGATGTTGCCGCTGCCGGCGTTTC 3963682

      86 ...LeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeuSe 101
            |||:::      |||   ||||||:::::::::::::|||   ||
  3963683CGGTGCCGACGTTGTTGTTGCCGGTATTCCCGAACCCGACGTTGG...AG 3963729

     101 rIleAlaTyrGlyArgCys 107
         |      |||   ||||||
  3963730TTGCCGGTATTGCCGCTGC 3963748

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-653-740-4

seq_documentation_block:
; Sequence 4, Application US/08653740
; Patent No. 5792850
;  GENERAL INFORMATION:
;    APPLICANT:  James W. Baumgartner
;    APPLICANT:  Donald C. Foster
;    APPLICANT:  Frank J. Grant
```

```
;    APPLICANT:  Cindy A. Sprecher
;    TITLE OF INVENTION:  HEMATOPOIETIC CYTOKINE RECEPTOR
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,740
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31,648
;      REFERENCE/DOCKET NUMBER:  95-31
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2049
US-08-653-740-4


alignment_scores:
            Quality:   78.50              Length:    125
              Ratio:  1.287                Gaps:      8
 Percent Similarity:  48.800   Percent Identity: 28.800

alignment_block:
 US-09-215-435-179 x US-08-653-740-4/rev   ..

 Align seg 1/1  to reverse of: US-08-653-740-4  from: 1  to: 2663

       4 ProGlnTrpLeu.................................LeuLe 9
         |||:::||||||                                 |||::
    2533 CCCAAGTGGCTCGGACTACAGGTGTGTGCCATCATGCCCAGCTAATTGTT 2484

       9 uLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySer.... 24
         :|||||||||:::|||:::|||            ::::::||||||
    2483 TTTGTTTTTGATGTTTTATTTT.........TGTAGTGAGGGGTCTTGCT 2443

      25 ..LeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIle 40
          ::::::   :::      :::|||||   |||
    2442 ATATTGCCCAGGCTGGTCTCAAACTCCTAGCTCTA.............. 2408

      41 ArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCy 57
             :::|||   :::|||   :::|||||||   |||      :::
    2407 ......AGTGATCCTTCCACCTTGGCCTGCCAAAGTGCTGGGATTACAGG 2364

      57 sGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
           ||||||||||||      |||   :::   ::::::|||   :::
    2363 CGTGAGCCACTGTGCT......GGGGTCCAAATTCAGTTCTGTTACAGCG 2320

      74 lnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThr 90
         :::   ||||||      |||      |||         :::::::::
    2319 GAACAGATTTTGGC......AGGCAGCTGCCACTC......TCAATTTCA 2282

      91 CysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCy 107
         |||||||:::                 |||      ::::::      ||
    2281 TGTATTTTC.................CTTTGCTCCACCTTCCATCCGTG 2250

     107 sGlnLysIleGlyArgGlnLysLeu 115
         |         |||   |||   |||
    2249 TCTTCCCCATGGGGGCCAAGCCCTT 2225

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-09-073-594-4

seq_documentation_block:
; Sequence 4, Application US/09073594
; Patent No. 5925735
;  GENERAL INFORMATION:
;    APPLICANT:  James W. Baumgartner
;    APPLICANT:  Donald C. Foster
;    APPLICANT:  Frank J. Grant
;    APPLICANT:  Cindy A. Sprecher
;    TITLE OF INVENTION:  HEMATOPOIETIC CYTOKINE RECEPTOR
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,594
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31,648
;      REFERENCE/DOCKET NUMBER:  95-31
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2049
 US-09-073-594-4


 alignment_scores:
             Quality:   78.50              Length:    125
               Ratio:  1.287                Gaps:      8
  Percent Similarity:  48.800   Percent Identity: 28.800

 alignment_block:
  US-09-215-435-179 x US-09-073-594-4/rev   ..

  Align seg 1/1  to reverse of: US-09-073-594-4  from: 1  to: 2663

        4 ProGlnTrpLeu.................................LeuLe 9
          |||:::||||||                                 |||::
     2533 CCCAAGTGGCTCGGACTACAGGTGTGTGCCATCATGCCCAGCTAATTGTT 2484

        9 uLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySer.... 24
          :|||||||||:::|||:::|||            ::::::||||||
     2483 TTTGTTTTTGATGTTTTATTTT.........TGTAGTGAGGGGTCTTGCT 2443
```

```
      25 ..LeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIle 40
         ::::::   :::      :::||||||   |||
    2442 ATATTGCCCAGGCTGGTCTCAAACTCCTAGCTCTA..............  2408

      41 ArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCy 57
            :::|||   :::|||   :::||||||   |||      :::
    2407 ......AGTGATCCTTCCACCTTGGCCTGCCAAAGTGCTGGGATTACAGG 2364

      57 sGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
          ||||||||||||      |||   :::   ::::::|||   :::
    2363 CGTGAGCCACTGTGCT......GGGGTCCAAATTCAGTTCTGTTACAGCG 2320

      74 lnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThr 90
         :::   ||||||      |||      |||         :::::::::
    2319 GAACAGATTTTGGC......AGGCAGCTGCCACTC......TCAATTTCA 2282

      91 CysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCy 107
         ||||||:::                 |||      ::::::      ||
    2281 TGTATTTTC.................CTTTGCTCCACCTTCCATCCGTG 2250

     107 sGlnLysIleGlyArgGlnLysLeu 115
         |         |||   |||   |||
    2249 TCTTCCCCATGGGGGCCAAGCCCTT 2225

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-275-925-4

seq_documentation_block:
; Sequence 4, Application US/09275925
; Patent No. 6080406
;  GENERAL INFORMATION:
;    APPLICANT:  James W. Baumgartner
;    APPLICANT:  Donald C. Foster
;    APPLICANT:  Frank J. Grant
;    APPLICANT:  Cindy A. Sprecher
;    TITLE OF INVENTION:  HEMATOPOIETIC CYTOKINE RECEPTOR
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/275,925
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31,648
;      REFERENCE/DOCKET NUMBER:  95-31
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2663 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  139..2049
US-09-275-925-4

alignment_scores:
            Quality:   78.50              Length:    125
              Ratio:  1.287                Gaps:      8
 Percent Similarity:  48.800   Percent Identity: 28.800

alignment_block:
 US-09-215-435-179 x US-09-275-925-4/rev   ..

 Align seg 1/1  to reverse of: US-09-275-925-4  from: 1  to: 2663

       4 ProGlnTrpLeu.................................LeuLe 9
         |||:::||||||                                 |||::
    2533 CCCAAGTGGCTCGGACTACAGGTGTGTGCCATCATGCCCAGCTAATTGTT 2484

       9 uLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySer.... 24
         :|||||||||:::|||:::|||           ::::::||||||
    2483 TTTGTTTTTGATGTTTTATTTT.........TGTAGTGAGGGGTCTTGCT 2443

      25 ..LeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIle 40
         ::::::   :::      :::||||||   |||
    2442 ATATTGCCCAGGCTGGTCTCAAACTCCTAGCTCTA..............  2408

      41 ArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCy 57
            :::|||   :::|||   :::||||||   |||      :::
    2407 ......AGTGATCCTTCCACCTTGGCCTGCCAAAGTGCTGGGATTACAGG 2364

      57 sGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrG 74
          ||||||||||||      |||   :::   ::::::|||   :::
    2363 CGTGAGCCACTGTGCT......GGGGTCCAAATTCAGTTCTGTTACAGCG 2320

      74 lnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThr 90
         :::   ||||||      |||      |||         :::::::::
    2319 GAACAGATTTTGGC......AGGCAGCTGCCACTC......TCAATTTCA 2282

      91 CysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCy 107
         ||||||:::                 |||      ::::::      ||
    2281 TGTATTTTC.................CTTTGCTCCACCTTCCATCCGTG 2250

     107 sGlnLysIleGlyArgGlnLysLeu 115
         |         |||   |||   |||
    2249 TCTTCCCCATGGGGGCCAAGCCCTT 2225

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-085-199B-3

seq_documentation_block:
; Sequence 3, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Larson, Marina T.
;       REGISTRATION NUMBER:  32038
;       REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (970) 668-2050
;       TELEFAX:  (970) 668-2052
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4796
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  no
;     ANTI-SENSE:  no
;     ORIGINAL SOURCE:
;       ORGANISM:  human
;     FEATURE:
;       OTHER INFORMATION:  cDNA for Huntington-interacting protein
US-09-085-199B-3


alignment_scores:
            Quality:   78.50              Length:    108
              Ratio:  1.353                Gaps:      5
 Percent Similarity:  53.704   Percent Identity: 26.852

alignment_block:
 US-09-215-435-179 x US-09-085-199B-3/rev   ..

 Align seg 1/1  to reverse of: US-09-085-199B-3  from: 1  to: 4796

      13 LeuPhePhePheLeuPheLeuLeuThrArgGlySerLeu..... 25
         |||:::|||   :::  ||||||  |||  ||||||
    1385 CTGTACTTCTCCTTTAGCTTGCTATATCGCTGTTCATTGGCTTGAGCTTT 1336

      26 .............................SerProThrLysTyrA 31
                                        |||   :::::
    1335 CCTTTCTATCTCAGACAGGCTCCGCTGAGCCTTCTCGGTGTCCTCCCGCT 1286

      31 snLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThr 47
         ||||||:::   ::::::|||:::  ||||||:::::   :::::
    1285 GCCTCCTGAGCTCGTCCAGTTCTGCCCGCAGGAATTCACAGTCGTCGGCC 1236

      48 GlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLy 64
         :::|||||||:::|||:::         |||         |||:::   ::
    1235 GCCTGCTGCCGCAGGTGC.........TGC.........TGCTCGGCCAG 1204

      64 sGlySerGluGlySerLeu.............CysGlnThrGlnValP 76
         :::::::::::::||||||              |||:::|||  .    :
    1203 ATCTGCTTCCAGCTCGCTGACGTGGCCCTTCAGCTGCAGCACAACCCGCT 1154

      76 hePheGlyGlnTyrArgAlaCysProCysLeuArgAsn...LeuThrCys 91
         :::::::   :::          |||     |||   |||
    1153 GGCTCTCAGTCTTCATGTTTTCTAGCTGTGCCTTCAATCCACTGATCTCT 1104

      92 IleTyrSerLysAsnGluLysTrp 99
         :::|||||||::::::   ||||||
    1103 CTGTATAGTCGCTCAATTAAGTGG 1080

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-142-078-63

seq_documentation_block:
; Sequence 63, Application US/09142078
; Patent No. 6172041
;  GENERAL INFORMATION:
;    APPLICANT:  McCabe, R. Tyler
;    APPLICANT:  Zhou, Li-Ming
;    APPLICANT:  Layer, Richard T.
;    TITLE OF INVENTION:  Use of Conantokins
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rothwell, Figg, Ernst & Kurz, p.c.
;      STREET:  555 Thirteenth Street, N.W., Suite 701-E
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/142,078
;      FILING DATE:  10-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO US97/12652
;      FILING DATE:  21-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/762,377
;      FILING DATE:  06-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/684,750
;      FILING DATE:  22-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  2314-135.A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-783-6040
;      TELEFAX:  202-783-6031
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  356 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (cDNA)
;    ORIGINAL SOURCE:
;      ORGANISM:  Conus caracteristicus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..279
US-09-142-078-63


alignment_scores:
            Quality:   77.50              Length:     91
              Ratio:  1.550                Gaps:      4
 Percent Similarity:  54.945   Percent Identity: 29.670

alignment_block:
 US-09-215-435-179 x US-09-142-078-63/rev   ..

 Align seg 1/1  to reverse of: US-09-142-078-63  from: 1  to: 356

      11 PheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySer...LeuSe 26
         |||      :::|||||||:::|||||||||   |||           ||
     298 TTTAGAACTATCTTTTTCATTTTCCTGCTTCTTCGAGTTCTCTAACAGTC 249

      26 rProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIle..... 40
         ||||   :::     ||||||:::            |||:::
     248 TCCGCAATCTCTCTATCTTCTTCGTAGCCTCTGGCCATGTTTCCTCGCTT 199

      41 .....ArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAsp 55
           ::::::::::::   |||:::  |||          |||
     198 TTTGAGAGTCCTCTTCCTCTGAGTCAACCTGT.....CCTTGC 161

      56 AsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGl 72
         :::|||:::::::   |||:::::::::   |||  |||:::  :::||
     160 CATTGTCGTCAGTGCTGCGGGCAGCGGATTTCTGCAGGAGGACAGGCTCA 111

      72 nThrGlnValPhePheGlyGlnTyrArg..................... 81
         |:::::::|||::::::   |||  :::
     110 GGTTTCAGTGCTGTGGCATCAGCCGAACGGCGTTCAGTCAGTGCGTCTCC 61
```

```
      82 ........AlaCysProCysLeu 86
                 ||||||||||||||||
      60 ATGAGCTAGCGTGCCCGTGCCTA 38

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-357-141-63

seq_documentation_block:
; Sequence 63, Application US/09357141
; Patent No. 6277825
; GENERAL INFORMATION:
;  APPLICANT:  Olivera, Baldomero M.
;  APPLICANT:  McIntosh, J. Michael
;  APPLICANT:  McCabe, R. Tyler
;  APPLICANT:  Layer, Richard T.
;  APPLICANT:  Zhou, Li-Ming
;  TITLE OF INVENTION:  Use of Conantokins for Treating Pain
;  FILE REFERENCE:  2314-171
;  CURRENT APPLICATION NUMBER: US/09/357,141
;  CURRENT FILING DATE:  1999-07-20
;  PRIOR APPLICATION NUMBER:  US 09/283,277
;  PRIOR FILING DATE:  1999-04-01
;  PRIOR APPLICATION NUMBER:  US 09/142,078
;  PRIOR FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  WO US97/12652
;  PRIOR FILING DATE:  1997-07-21
;  PRIOR APPLICATION NUMBER:  US 08/762,377
;  PRIOR FILING DATE:  1996-12-06
;  PRIOR APPLICATION NUMBER:  US 08/684,750
;  PRIOR FILING DATE:  1996-07-22
;  NUMBER OF SEQ ID NOS:  71
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 63
;   LENGTH: 356
;   TYPE: DNA
;   ORGANISM: Conus caracteristicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(279)
US-09-357-141-63


alignment_scores:
            Quality:   77.50            Length:     91
              Ratio:  1.550              Gaps:      4
 Percent Similarity:  54.945   Percent Identity: 29.670

alignment_block:
 US-09-215-435-179 x US-09-357-141-63/rev   ..

 Align seg 1/1  to reverse of: US-09-357-141-63  from: 1  to: 356

      11 PheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySer...LeuSe 26
         |||      :::||||||:::|||||||||   |||            ||
     298 TTTAGAACTATCTTTTTCATTTTCCTGCTTCTTCGAGTTCTCTAACAGTC 249

      26 rProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIle..... 40
         ||||   :::      ||||||:::          |||:::
     248 TCCGCAATCTCTCTATCTTCTTCGTAGCCTCTGGCCATGTTTCCTCGCTT 199

      41 .....ArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAsp 55
           ::::::::::::   |||:::   |||            |||
     198 TTTGAGAGTCCTCTTCCTCTGAGTCAACCTGT...........CCTTGC 161

      56 AsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGl 72
         :::|||::::::   |||:::::::::   |||   |||:::   :::||
     160 CATTGTCGTCAGTGCTGCGGGCAGCGGATTTCTGCAGGAGGACAGGCTCA 111

      72 nThrGlnValPhePheGlyGlnTyrArg..................... 81
         |:::::::|||::::::   |||   :::
     110 GGTTTCAGTGCTGTGGCATCAGCCGAACGGCGTTCAGTCAGTGCGTCTCC 61

      82 ........AlaCysProCysLeu 86
                 ||||||||||||||||
      60 ATGAGCTAGCGTGCCCGTGCCTA 38

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-11

seq_documentation_block:
; Sequence 11, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  136..2543
US-07-862-021B-11


alignment_scores:
            Quality:   76.00            Length:    101
              Ratio:  1.407              Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179 x US-07-862-021B-11   ..

 Align seg 1/1  to: US-07-862-021B-11  from: 1  to: 3226

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||      |||:::|||||   |||::::::|||:::   :::   ||
    1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

      50 sGlnArgAlaPro............AspAsnCysGluSerHisCysAlaG 63
         |      :::               |||:::|||::::::   |||:::
    1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

      63 lu..............................LysGlySerGluGly 68
                                                 ::::::|||
    1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896
```

```
     69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
        |||:::|||:::::::::::::      ::::::            |||
   1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

     85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
          :::::::::   |||:::  |||      ::::::|||
   1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

    100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
         |||::::::   |||:::  ::::::     ||||||::::::
   1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

    117 Lys 117
        |||
   2044 AAA 2046
```

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-11

```
seq_documentation_block:
; Sequence 11, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  136..2543
US-08-313-288B-11


alignment_scores:
            Quality:   76.00              Length:    101
              Ratio:  1.407                Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179 x US-08-313-288B-11   ..

 Align seg 1/1  to: US-08-313-288B-11  from: 1  to: 3226
```

```
     34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
        |||      |||:::||||||   |||::::::|||:::   :::   ||
   1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

     50 sGlnArgAlaPro.............AspAsnCysGluSerHisCysAlaG 63
        |      :::                |||:::|||::::::   |||:::
   1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

     63 lu................................LysGlySerGluGly 68
                                                 ::::::|||
   1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896

     69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
        |||:::|||:::::::::::::      ::::::            |||
   1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

     85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
          :::::::::   |||:::  |||      ::::::|||
   1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

    100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
         |||::::::   |||:::  ::::::     ||||||::::::
   1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

    117 Lys 117
        |||
   2044 AAA 2046
```

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-11

```
seq_documentation_block:
; Sequence 11, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  136..2543
PCT-US93-03164-11


alignment_scores:
            Quality:   76.00              Length:    101
              Ratio:  1.407                Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179 x PCT-US93-03164-11   ..

 Align seg 1/1  to: PCT-US93-03164-11  from: 1  to: 3226

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||      |||:::||||||   |||:::::::|||:::   :::   ||
    1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

      50 sGlnArgAlaPro...........AspAsnCysGluSerHisCysAlaG 63
         |      :::               |||:::|||::::::   |||:::
    1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

      63 lu.............................LysGlySerGluGly 68
                                        ::::::|||
    1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896

      69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
         |||:::|||:::::::::::      ::::::             |||
    1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

      85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
           :::::::::   |||:::   |||      :::::::|||
    1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

     100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
         |||::::::   |||:::   ::::::      ||||||::::::
    1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

     117 Lys 117
         |||
    2044 AAA 2046

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-3

seq_documentation_block:
; Sequence 3, Application US/08597545
; Patent No. 5580738
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,545
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Human Dlk
US-08-597-545-3


alignment_scores:
            Quality:   75.00              Length:     61
              Ratio:  2.143                Gaps:      5
 Percent Similarity:  57.377   Percent Identity: 34.426

alignment_block:
 US-09-215-435-179 x US-08-597-545-3   ..

 Align seg 1/1  to: US-08-597-545-3  from: 1  to: 1556

      35 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 51
         |||   ::::::|||:::   :::      |||      |||   |||
     330 CTTTGTGACCAGTGCGTGACCTCTCCCGGCTGCCTTCACGGACTCTGTGG 379

      51 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 68
          :::   |||   :::|||         |||::::::   |||   :::|
     380 AGAA...CCCGGGCAGTGC......ATTTGCACCGAC...GGCTGGGACG 417

      68 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 83
         ||:::||||||:::   :::|||               |||||||||
     418 GGGAGCTCTGTGATAGAGATGTT..............CGGGCCTGCTCC 452

      84 ......ProCysLeuArgAsnLeuThrCysIle 92
               ||||||   :::|||   ||||||:::
     453 TCGGCCCCCTGTGCCAACAACGGGACCTGCGTG 485

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-3

seq_documentation_block:
; Sequence 3, Application US/08457135
; Patent No. 5644031
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,135
;      FILING DATE:  01-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
```

```
;      REFERENCE/DOCKET NUMBER:  40399/304/NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Human Dlk
US-08-457-135-3


alignment_scores:
            Quality:   75.00              Length:     61
              Ratio:  2.143                Gaps:      5
 Percent Similarity:  57.377   Percent Identity: 34.426

alignment_block:
 US-09-215-435-179 x US-08-457-135-3   ..

 Align seg 1/1  to: US-08-457-135-3  from: 1  to: 1556

      35 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 51
         |||   ::::::|||:::   :::      |||      |||   |||
     330 CTTTGTGACCAGTGCGTGACCTCTCCCGGCTGCCTTCACGGACTCTGTGG 379

      51 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 68
          :::   |||   :::|||        |||::::::   |||   :::|
     380 AGAA...CCCGGGCAGTGC......ATTTGCACCGAC...GGCTGGGACG 417

      68 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 83
         ||:::|||||:::   :::|||              |||||||||
     418 GGGAGCTCTGTGATAGAGATGTT.............CGGGCCTGCTCC 452

      84 ......ProCysLeuArgAsnLeuThrCysIle 92
              ||||||   :::|||   ||||||:::
     453 TCGGCCCCCTGTGCCAACAACGGGACCTGCGTG 485

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-882-046-3

seq_documentation_block:
; Sequence 3, Application US/08882046
; Patent No. 6136952
;  GENERAL INFORMATION:
;    APPLICANT:  Li, Linheng
;    APPLICANT:  Hood, Leroy
;    APPLICANT:  Krantz, Ian D.
;    APPLICANT:  Spinner, Nancy B.
;    TITLE OF INVENTION:  Human Jagged Polypeptide, Encoding
;    TITLE OF INVENTION:  Nucleic Acids and Methods of Use
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/882,046
;      FILING DATE:  25-JUN-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UW 2637
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4315 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16..3460
US-08-882-046-3


alignment_scores:
            Quality:   75.00              Length:     99
              Ratio:  1.531                Gaps:      4
 Percent Similarity:  49.495   Percent Identity: 23.232

alignment_block:
 US-09-215-435-179 x US-08-882-046-3   ..

 Align seg 1/1  to: US-08-882-046-3  from: 1  to: 4315

      31 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 47
         |||      :::   :::::::::|||      |||      ::::::
     892 AACTGTGAGAAGGCTGAGCACGCCTGCACCTCCAACCCGTGTGCCAACGG 941

      47 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 64
          |||   |||::::::::::|||::::::   |||   ||||||      :
     942 GGGCTCTTGCCATGAGGTGCCGTCCGGCTTCGAATGCCACTGC...CCAT 988

      64 ysGlySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyr 80
         ::|||   :::|||      |||      ::::::        :::
     989 CGGGCTGGAGCGGGCCCACCTGTGCCCTTGACATC......GATGAGTGT 1032

      81 ArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn.. 96
            :::   ||||||      :::   ||||||:::   :::   :::
    1033 GCTTCGAACCCGTGTGCGGCCGGTGGCACCTGTGTGGACCAGGTGGACGG 1082

      97 ..................GluLysTrpLeuSerIle..... 102
                            |||:::|||::::::
    1083 CTTTGAGTGCATCTGCCCCGAGCAGTGGGTGGGGGCCACCTGCCAGCTGG 1132

     103 ..........AlaTyrGlyArgCysGlnLysIleGlyArgGlnLys 114
                  :::   |||:::||||||      |||      |||
    1133 ACGTCAACGACTGTCGCGGGCAGTGTCAGCATGGGGGCACCTGCAAG 1179

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-181-706-1

seq_documentation_block:
; Sequence 1, Application US/09181706
; Patent No. 6130068
;  GENERAL INFORMATION:
;    APPLICANT:  Melanie K. Spriggs, Michael R. Comeau,
;    APPLICANT:  Robert F. DuBose, Richard S. Johnson
;    TITLE OF INVENTION:  VIRAL ENCODED SEMAPHORIN PROTEIN
;    TITLE OF INVENTION:  RECEPTOR DNA AND POLYPEPTIDES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janis C. Henry
;      STREET:  51 University St.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  US
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/181,706
;       FILING DATE:  October 28, 1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/958,598 (converted to a
;       APPLICATION NUMBER:  Provisional, see below)
;       FILING DATE:  October 28, 1997
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  --to be assigned-- (USSN 08/958,598
;       APPLICATION NUMBER:  conversion to Provisional application)
;       FILING DATE:  October 26, 1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Henry, Janis C
;       REGISTRATION NUMBER:  34,347
;       REFERENCE/DOCKET NUMBER:  2631-A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206)470-4189
;       TELEFAX:  (206)233-0644
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4707 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..4707
US-09-181-706-1


alignment_scores:
            Quality:   75.00              Length:     78
              Ratio:  1.705                Gaps:      7
 Percent Similarity:  56.410   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179 x US-09-181-706-1   ..

 Align seg 1/1  to: US-09-181-706-1  from: 1  to: 4707

      56 AsnCysGlu..............SerHisCysAlaGluLysGlySerGl 67
         :::|||||||              |||:::      |||:::|||:::|| 
    1296 TCCTGTGAAGAATATCTACATTTATCTAACAGCTGGGAAAGAGGTGAGGA 1345

      67 uGlySerLeuCysGlnThrGln.........ValPhePheGlyGlnTyr. 80
         |:::|||   |||:::   |||         ::::::||||||   :::  
    1346 GAATTCGTGTTGCAAACTGCAATAAACATAAATCCTGTTCGGAGTGTTTA 1395

      81 ....................ArgAla.CysProCysLeuArgAsnLeuTh 90
                             |||    |||      |||::::::   ||  
    1396 ACAGCCACAGACCCTCACTGCGGTTGGTGCCATTCGCTACAAAGGTGCAC 1445

      90 r............CysIleTyrSerLysAsn...GluLysTrpLeuSerI 102
         |            |||:::::::|||:::|||   |||:::||||||:::| 
    1446 TTTTCAAGGAGATTGTGTACATTCAGAGAACTTAGAAAACTGGCTGGATA 1495

     102 leAlaTyrGly......ArgCysGlnLysIle 110
         ||:::   |||      :::|||   |||||| 
    1496 TTTCGTCTGGAGCAAAAAAGTGCCCTAAAATT 1527

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-458-791-1

seq_documentation_block:
; Sequence 1, Application US/09458791
; Patent No. 6174689
;   GENERAL INFORMATION:
;       APPLICANT: Spriggs, Melanie
;       TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
;                           RECEPTOR DNA AND POLYPEPTIDES
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Janis C. Henry
;            STREET: 51 University St.
;            CITY: Seattle
;            STATE: WA
;            COUNTRY: US
;            ZIP: 98101
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: MS-DOS/Windows 95
;            SOFTWARE: Word for Windows 95, 7.0a
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/458,791
;            FILING DATE: 10-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/958,598
;            FILING DATE: 28-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Henry, Janis C
;            REGISTRATION NUMBER: 34,347
;            REFERENCE/DOCKET NUMBER: 2631
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206)470-4189
;            TELEFAX: (206)233-0644
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 4707 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..4707
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-458-791-1


alignment_scores:
            Quality:   75.00              Length:     78
              Ratio:  1.705                Gaps:      7
 Percent Similarity:  56.410   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179 x US-09-458-791-1   ..

 Align seg 1/1  to: US-09-458-791-1  from: 1  to: 4707

      56 AsnCysGlu..............SerHisCysAlaGluLysGlySerGl 67
         :::|||||||              |||:::      |||:::|||:::|| 
    1296 TCCTGTGAAGAATATCTACATTTATCTAACAGCTGGGAAAGAGGTGAGGA 1345

      67 uGlySerLeuCysGlnThrGln.........ValPhePheGlyGlnTyr. 80
         |:::|||   |||:::   |||         ::::::||||||   :::  
    1346 GAATTCGTGTTGCAAACTGCAATAAACATAAATCCTGTTCGGAGTGTTTA 1395

      81 ....................ArgAla.CysProCysLeuArgAsnLeuTh 90
                             |||    |||      |||::::::   ||  
    1396 ACAGCCACAGACCCTCACTGCGGTTGGTGCCATTCGCTACAAAGGTGCAC 1445

      90 r............CysIleTyrSerLysAsn...GluLysTrpLeuSerI 102
         |            |||::::::|||:::|||   |||:::||||||:::| 
    1446 TTTTCAAGGAGATTGTGTACATTCAGAGAACTTAGAAAACTGGCTGGATA 1495

     102 leAlaTyrGly......ArgCysGlnLysIle 110
```

```
     ||:::   |||      :::|||   ||||||
1496 TTTCGTCTGGAGCAAAAAAGTGCCCTAAAATT 1527
```

```
OM of: US-09-215-435-179 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78 +  666.00 1253.41  1.8e-60   542
/cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61 +  666.00 1252.98  1.9e-60   565
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73 +  666.00 1251.08  2.4e-60   680
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73 +  666.00 1251.08  2.4e-60   680
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817 +  661.00 1244.08  5.9e-60   533
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817 +  661.00 1244.08  5.9e-60   53
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44 +  619.00 1166.77  1.2e-55   421
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481-520 +  602.00 1132.98  9.2e-54   488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481A-520 +  602.00 1132.98  9.2e-54   488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481B-520 +  602.00 1132.98  9.2e-54   488
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-48 +  497.00  934.19  1.1e-42   463
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-694 +  497.00  933.86  1.1e-42   478
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-31337-1 +  497.00  932.94  1.3e-42   523
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-443-634-1 +  497.00  932.94  1.3e-42   523  !
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-713-067-1 +  497.00  932.94  1.3e-42   523
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-731-872-89 +  383.50  717.65  1.2e-30   514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5913 +  383.50  717.65  1.2e-30   514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-876-997-89 +  383.50  717.65  1.2e-30   514
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-1374 +  383.50  717.65  1.2e-30   51
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-108 +  383.50  717.65  1.2e-30   514
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-108 +  383.50  717.65  1.2e-30   514
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5913 +  383.50  717.65  1.2e-30   51
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-992-332-1466 +  383.00  723.58  5.8e-31   263
/cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-033-364-14 +  383.00  723.58  5.8e-31   263
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-192 +  346.00  586.81  2.4e-23   171
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-219 +  346.00  586.81  2.4e-23   171
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-409-86528 +  188.00  347.96  4.9e-10   4
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-933-524-86528 +  188.00  347.96  4.9e-10   44
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-274-861-9011 +  135.50  249.49  0.0001   397
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-915-738-9011 +  135.50  249.49  0.0001   397
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-284-697A-15 +  135.50  246.66  0.0002   523
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-289898 +  135.50  245.69  0.0002
/cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-289898 +  135.50  245.69  0.0002
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-213-360-121 +  135.50  245.43  0.0003   590
/cgn1_7/ptodata/1/pna/US6032_COMB.seq:US-60-324-185-6544 +  135.50  245.43  0.0003   59
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-7251 +  135.50  245.25  0.0003   60
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-647-801-16 +  135.50  244.34  0.0003   656
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-05989-156 +  135.50  244.18  0.0003   666
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-925-297-156 +  135.50  244.18  0.0003   666
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-8817 +  135.50  243.03  0.0003   74
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-122 +  135.00  244.09  0.0003
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-821-710-1539 +  134.00  242.42  0.0004
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-14399 +  131.00  229.88  0.0018
/cgn1_7/ptodata/1/pna/US6032_COMB.seq:US-60-324-185-6539 +  131.00  229.25  0.0020
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-705-426-361 +  131.00  229.22  0.0020

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78

seq_documentation_block:
; Sequence 78, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 78
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 16..378
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 16..84
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.8
;   OTHER INFORMATION: seq FLLFFFLFLLLTRG/SL
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 502..507
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 522..542
US-09-215-435-78


alignment_scores:
            Quality:  666.00              Length:    121
              Ratio:  5.504                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-09-215-435-78   ..

 Align seg 1/1  to: US-09-215-435-78  from: 1  to: 542

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      66 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATATAACCTTTTGG 115

      34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     116 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 165

      51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  166 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 215

   67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  216 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTC 265

   84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  266 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 315

  101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  316 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 365

  117 sLysMetPhePhe 121
      ||||||||||||
  366 GAAAATGTTCTTC 378

seq_name: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61

seq_documentation_block:
; Sequence 61, Application US/60096116
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  95
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/096,116
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.038PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  565 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  16..84
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLTRG/SL
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  502..507
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  529..542
;      IDENTIFICATION METHOD:  blastn2
;    FEATURE:
;      NAME/KEY:  vrt
;      LOCATION:  complement(33..74)
;      IDENTIFICATION METHOD:  fasta
;      OTHER INFORMATION:  identity 88
;      OTHER INFORMATION:  region 21722..21764
;      OTHER INFORMATION:  id AF048728
;    FEATURE:
;      NAME/KEY:  vrt
;      LOCATION:  511..545
;      IDENTIFICATION METHOD:  fasta
;      OTHER INFORMATION:  identity 83
;      OTHER INFORMATION:  region 4627..4663
;      OTHER INFORMATION:  id AF048728
US-60-096-116-61


alignment_scores:
           Quality:  666.00              Length:    121
             Ratio:  5.504                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-60-096-116-61   ..

 Align seg 1/1  to: US-60-096-116-61  from: 1  to: 565

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      66 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATATAACCTTTTGG 115

      34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     116 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 165

      51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     166 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 215

      67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     216 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTC 265

      84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     266 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 315

     101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     316 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 365

     117 sLysMetPhePhe 121
         ||||||||||||
     366 GAAAATGTTCTTC 378

 seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73

 seq_documentation_block:
 ; Sequence 73, Application PC/TUS9917130
 ; GENERAL INFORMATION:
 ;  APPLICANT: Human Genome Sciences, Inc.
 ;  TITLE OF INVENTION: 98 Human Secreted Proteins
 ;  FILE REFERENCE: PZ031.PCT
 ;  CURRENT APPLICATION NUMBER: PCT/US99/17130
 ;  CURRENT FILING DATE: 1999-07-29
 ;  EARLIER APPLICATION NUMBER: 60/094,657
 ;  EARLIER FILING DATE: 1998-07-30
 ;  EARLIER APPLICATION NUMBER: 60/095,486
```

```
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-17130-73


alignment_scores:
            Quality:  666.00            Length:    121
              Ratio:  5.504               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x PCT-US99-17130-73   ..

 Align seg 1/1  to: PCT-US99-17130-73  from: 1  to: 680

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     183 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 232

      34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     233 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 282

      51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     283 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 332

      67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     333 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAGTATAGAGCGTGTC 382

      84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     383 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 432

     101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     433 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 482

     117 sLysMetPhePhe 121
         ||||||||||||
     483 GAAAATGTTCTTC 495

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73

seq_documentation_block:
; Sequence 73, Application US/09489847
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031P1
;  CURRENT APPLICATION NUMBER: US/09/489,847
;  CURRENT FILING DATE: 2000-01-24
;  EARLIER APPLICATION NUMBER: PCT/US99/17130
;  EARLIER FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 376
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-489-847-73


alignment_scores:
            Quality:  666.00            Length:    121
              Ratio:  5.504               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-09-489-847-73   ..

 Align seg 1/1  to: US-09-489-847-73  from: 1  to: 680

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     183 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 232
```

```
     34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    233 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 282

     51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    283 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 332

     67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    333 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAGTATAGAGCGTGTC 382

     84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    383 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 432

    101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    433 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 482

    117 sLysMetPhePhe 121
        |||||||||||||
    483 GAAAATGTTCTTC 495

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817

seq_documentation_block:
; Sequence 4817, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 4817
;   LENGTH: 533
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 99..461
;   NAME/KEY: sig_peptide
;   LOCATION: 99..167
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.80000019073486
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-09-834-366-4817


alignment_scores:
            Quality:  661.00            Length:    121
              Ratio:  5.508               Gaps:      0
 Percent Similarity:  99.174   Percent Identity: 99.174

alignment_block:
 US-09-215-435-179 x US-09-834-366-4817   ..

 Align seg 1/1  to: US-09-834-366-4817  from: 1  to: 533

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
```

```
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     149 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 198

      34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     199 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 248

      51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     249 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 298

      67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
         |||||||   ||||||||||||||||||||||||||||||||||||||||
     299 GGGCAGTCKGTGTCAAACGCAGGTGTTCTTTGGCCARTATAGAGCGTGTC 348

      84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     349 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 398

     101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     399 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 448

     117 sLysMetPhePhe 121
         |||||||||||||
     449 RAAAATGTTCTTC 461

seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817

seq_documentation_block:
; Sequence 4817, Application US/60197873
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
;  FILE REFERENCE:  81.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/197,873
;  CURRENT FILING DATE:  2000-04-18
;  NUMBER OF SEQ ID NOS:  52153
;  SOFTWARE:  Patent.pm
; SEQ ID NO 4817
;   LENGTH: 533
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 99..461
;   NAME/KEY: sig_peptide
;   LOCATION: 99..167
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.80000019073486
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-60-197-873-4817


alignment_scores:
            Quality:  661.00            Length:    121
              Ratio:  5.508               Gaps:      0
 Percent Similarity:  99.174   Percent Identity: 99.174

alignment_block:
 US-09-215-435-179 x US-60-197-873-4817   ..

 Align seg 1/1  to: US-60-197-873-4817  from: 1  to: 533

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      149 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 198

       34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      199 AGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTGC 248

       51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      249 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 298

       67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
          |||||||   ||||||||||||||||||||||||||||||||||||||||
      299 GGGCAGTCKGTGTCAAACGCAGGTGTTCTTTGGCCARTATAGAGCGTGTC 348

       84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      349 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 398

      101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      399 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAA 448

      117 sLysMetPhePhe 121
          |||||||||||||
      449 RAAAATGTTCTTC 461

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44

seq_documentation_block:
; Sequence 44, Application US/08905279
;  GENERAL INFORMATION:
;    APPLICANT:  Genset SA
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,279
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.011A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  421 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  62..130
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLLTRG/SL
US-08-905-279-44


alignment_scores:
            Quality:  619.00              Length:    118
              Ratio:  5.383                 Gaps:      0
 Percent Similarity:  97.458   Percent Identity: 95.763

alignment_block:
 US-09-215-435-179 x US-08-905-279-44   ..

 Align seq 1/1  to: US-08-905-279-44  from: 1  to: 421

        1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       62 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 111

       17 uPheLeuLeuThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuG 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      112 CTTCCTCCTCACCAGGGGCTCACTTTCTCCAACAAAATACAACCTTTTGG 161

       34 luLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCys 50
          ||||||||||||:::          ||||||||||||||||||||||||||
      162 AGCTCAAGGAGKSKKGCATSGGGAACCAGGACTGCGAGACTGGCTGCTGC 211

       51 GlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGl 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      212 CAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCGA 261

       67 uGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysP 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      262 GGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGTC 311

       84 roCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      312 CCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCTT 361

      101 SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaLy 117
          ||||||||||||||||||||||||||||||||||||||||||||||||::
      362 AGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTAG 411

      117 sLys 118
          :|||
      412 RAAA 415

seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481-520

seq_documentation_block:
; Sequence 520, Application US/08196481
;  GENERAL INFORMATION:
;    APPLICANT:  Haseltine, William
;    APPLICANT:  Rosen, Craig
;    APPLICANT:  Ruben, Steve
;    APPLICANT:  Dillon, Patrick
;    APPLICANT:  Li, Haodong
;    TITLE OF INVENTION:  Human Genes, Sequences, and Expression
;    TITLE OF INVENTION:  Products
;    NUMBER OF SEQUENCES:  1719
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart, &
;      ADDRESSEE:  Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 5.0
;      SOFTWARE:  ASCII Text
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/196,481
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  325800-92
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 994-1700
;      TELEFAX:  (201) 994-1744
;  INFORMATION FOR SEQ ID NO:  520:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-196-481-520


alignment_scores:
            Quality:  602.00              Length:    112
              Ratio:  5.423                 Gaps:      0
 Percent Similarity:  99.107   Percent Identity: 97.321

alignment_block:
 US-09-215-435-179 x US-08-196-481-520   ..

 Align seg 1/1  to: US-08-196-481-520  from: 1  to: 488

      10 LeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSe 26
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      14 CTGTTCCTTCTCTTCTTCTTTCTCTTCCTCCTCACCAGGGGCTCACTTTC 63

      26 rProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAsnG 43
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 TCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCATCCGGAACC 113

      43 lnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGluSer 59
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 AGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCG 163

      60 HisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnValPh 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 CACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGTT 213

      76 ePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysIleT 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 CTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGACTTGTATAT 263

      93 yrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCysGlnLys 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 ATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTTGTCAGAAA 313

     110 IleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
         |||   ||||||||||||:::|||:::|||||||||
     314 ATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349

 seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481A-520

 seq_documentation_block:
 ; Sequence 520, Application US/08196481A
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Haseltine, William A.
 ;    APPLICANT:  Rosen, Craig A.
 ;    APPLICANT:  Ruben, Steve
 ;    APPLICANT:  Dillon, Patrick J.
 ;    APPLICANT:  Li, Haodong
 ;    TITLE OF INVENTION:  Human Genes, Sequences and Expression
 ;    NUMBER OF SEQUENCES:  1721
```

Products

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/196,481A
;      FILING DATE:  February 15, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenley K. Hoover
;      REGISTRATION NUMBER:  40,302
;      REFERENCE/DOCKET NUMBER:  PO-07
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8439
;  INFORMATION FOR SEQ ID NO:  520:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-196-481A-520


alignment_scores:
            Quality:  602.00              Length:    112
              Ratio:  5.423                 Gaps:      0
 Percent Similarity:  99.107   Percent Identity: 97.321

alignment_block:
 US-09-215-435-179 x US-08-196-481A-520   ..

 Align seg 1/1  to: US-08-196-481A-520  from: 1  to: 488

      10 LeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSe 26
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      14 CTGTTCCTTCTCTTCTTCTTTCTCTTCCTCCTCACCAGGGGCTCACTTTC 63

      26 rProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAsnG 43
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 TCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCATCCGGAACC 113

      43 lnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGluSer 59
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 AGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCG 163

      60 HisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnValPh 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 CACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGTT 213

      76 ePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysIleT 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 CTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGACTTGTATAT 263

      93 yrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCysGlnLys 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 ATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTTGTCAGAAA 313

     110 IleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
         |||   ||||||||||||:::|||:::|||||||||
     314 ATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349
```

```
seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481B-520

seq_documentation_block:
; Sequence 520, Application US/08196481B
; GENERAL INFORMATION:
;  APPLICANT:  Rosen, et. al.
;  TITLE OF INVENTION:  Human Genes, Sequences, and Expression Products 7
;  FILE REFERENCE:  P007
;  CURRENT APPLICATION NUMBER: US/08/196,481B
;  CURRENT FILING DATE:  1994-02-15
;  NUMBER OF SEQ ID NOS:  1721
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 520
;   LENGTH: 488
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (220)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (317)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (354)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (356)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (358)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (366)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (370)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (378)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (425)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (432)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (433)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (434)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (437)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (439)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (440)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (449)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (455)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (457)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (458)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (463)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (482)
;   OTHER INFORMATION: n equals a,t,g, or c
US-08-196-481B-520


alignment_scores:
            Quality:  602.00              Length:    112
              Ratio:  5.423                 Gaps:      0
 Percent Similarity:  99.107   Percent Identity: 97.321

alignment_block:
 US-09-215-435-179 x US-08-196-481B-520   ..

 Align seg 1/1  to: US-08-196-481B-520  from: 1  to: 488

      10 LeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSe 26
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      14 CTGTTCCTTCTCTTCTTCTTTCTCTTCCTCCTCACCAGGGGCTCACTTTC 63

      26 rProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAsnG 43
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 TCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCATCCGGAACC 113

      43 lnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGluSer 59
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 AGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCG 163

      60 HisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnValPh 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 CACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGTT 213

      76 ePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysIleT 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 CTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGACTTGTATAT 263

      93 yrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCysGlnLys 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 ATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTTGTCAGAAA 313

     110 IleGlyArgGlnLysLeuAlaLysLysMetPhePhe 121
         |||   ||||||||||||:::|||:::|||||||||
     314 ATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-48

seq_documentation_block:
; Sequence 48, Application US/60184698
; GENERAL INFORMATION:
;  APPLICANT: Hodgson, David M.                    Jones, Anissa L.
;  APPLICANT: Lincoln, Stephen E.                  Yu, Jimmy Y.
;  APPLICANT: Russo, Frank D.                      Greenawalt, Lila B.
;  APPLICANT: Spiro, Peter A.                      Panzer, Scott R.
;  APPLICANT: Banville, Steve C.                   Roseberry, Ann M.
;  APPLICANT: Bratcher, Shawn R.                   Wright, Rachel J.
;  APPLICANT: Dufour, Gerard E.                    Chen, Wensheng
;  APPLICANT: Cohen, Howard J.                     Liu, Tommy
;  APPLICANT: Rosen, Bruce                         Yap, Pierre E.
;  APPLICANT: Shah, Purvi                          Stockdreher, Theresa K.
;  APPLICANT: Chalup, Michael S.                   Amshey, Stefan
;  APPLICANT: Hillman, Jennifer L.                 Fong, Willy Tuen
;  TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
;  FILE REFERENCE: PT-0115 P
;  CURRENT APPLICATION NUMBER: US/60/184,698
;  CURRENT FILING DATE: 2000-02-24
;  NUMBER OF SEQ ID NOS: 1132
;  SOFTWARE: PERL Program
; SEQ ID NO 48
;   LENGTH: 463
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 474641.1
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 347-368
;   OTHER INFORMATION: a, t, c, g, or other
US-60-184-698-48


alignment_scores:
            Quality:  497.00            Length:     88
              Ratio:  5.648               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-60-184-698-48    ..

 Align seg 1/1  to: US-60-184-698-48  from: 1  to: 463

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      61 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 110

      50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 160

      67 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 83
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     161 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 210

      84 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 260

     100 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 310

     117 ysLysMetPhePhe 121
         |||||||||||||
     311 AGAAAATGTTCTTC 324

 seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-694

 seq_documentation_block:
 ; Sequence 694, Application US/60184698
 ; GENERAL INFORMATION:
 ;  APPLICANT: Hodgson, David M.
 ;  APPLICANT: Lincoln, Stephen E.
 ;  APPLICANT: Russo, Frank D.
 ;  APPLICANT: Spiro, Peter A.
 ;  APPLICANT: Banville, Steve C.
 ;  APPLICANT: Bratcher, Shawn R.
 ;  APPLICANT: Dufour, Gerard E.
 ;  APPLICANT: Cohen, Howard J.
 ;  APPLICANT: Rosen, Bruce
 ;  APPLICANT: Shah, Purvi
 ;  APPLICANT: Chalup, Michael S.
 ;  APPLICANT: Hillman, Jennifer L.
                    Jones, Anissa L.
                Yu, Jimmy Y.
                    Greenawalt, Lila B.
                    Panzer, Scott R.
                Roseberry, Ann M.
                Wright, Rachel J.
                    Chen, Wensheng
                    Liu, Tommy
                Yap, Pierre E.
                    Stockdreher, Theresa K.
                Amshey, Stefan
                    Fong, Willy Tuen
 ;  TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
 ;  FILE REFERENCE: PT-0115 P
 ;  CURRENT APPLICATION NUMBER: US/60/184,698
 ;  CURRENT FILING DATE: 2000-02-24
 ;  NUMBER OF SEQ ID NOS: 1132
 ;  SOFTWARE: PERL Program
 ; SEQ ID NO 694
 ;   LENGTH: 478
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 474641.1.i
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 347-368
;   OTHER INFORMATION: a, t, c, g, or other
US-60-184-698-694


alignment_scores:
            Quality:  497.00            Length:     88
              Ratio:  5.648               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-60-184-698-694    ..

 Align seg 1/1  to: US-60-184-698-694  from: 1  to: 478

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      61 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 110

      50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 160

      67 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 83
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     161 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 210

      84 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 260

     100 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 310

     117 ysLysMetPhePhe 121
         |||||||||||||
     311 AGAAAATGTTCTTC 324

 seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-31337-1

 seq_documentation_block:
 ; Sequence 1, Application PC/TUS0031337
 ; GENERAL INFORMATION:
 ;  APPLICANT:  ZymoGenetics, Inc.
 ;  TITLE OF INVENTION:  EDUCATIONAL KIT AND METHOD USING
 ;  TITLE OF INVENTION:  COLIPASE-LIKE PROTEIN-1
 ;  FILE REFERENCE:  99-85PC
 ;  CURRENT APPLICATION NUMBER: PCT/US00/31337
 ;  CURRENT FILING DATE:  2000-11-15
 ;  PRIOR APPLICATION NUMBER:  60/166,626
 ;  PRIOR FILING DATE:  1999-11-19
 ;  NUMBER OF SEQ ID NOS:  11
 ;  SOFTWARE:  FastSEQ for Windows Version 3.0
 ; SEQ ID NO 1
 ;   LENGTH: 523
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: (1)...(354)
 PCT-US00-31337-1


 alignment_scores:
             Quality:  497.00            Length:     88
               Ratio:  5.648               Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-215-435-179 x PCT-US00-31337-1   ..

 Align seg 1/1  to: PCT-US00-31337-1  from: 1  to: 523

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

      50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
         |||||||||||||||||||||||||||||||||||||||||||||||||
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      67 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 83
         |||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      84 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

     100 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 117
         |||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

     117 ysLysMetPhePhe 121
         |||||||||||||
     341 AGAAAATGTTCTTC 354

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-443-634-1

seq_documentation_block:
; Sequence 1, Application US/09443634
; GENERAL INFORMATION:
;  APPLICANT: Holloway, James L.
;  TITLE OF INVENTION: Colipase-Like Protein - 1
;  FILE REFERENCE: 99-85
;  CURRENT APPLICATION NUMBER: US/09/443,634
;  CURRENT FILING DATE: 1999-11-19
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 523
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(354)
US-09-443-634-1


alignment_scores:
            Quality:  497.00             Length:     88
              Ratio:  5.648                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-09-443-634-1   ..

 Align seg 1/1  to: US-09-443-634-1  from: 1  to: 523

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

      50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
         |||||||||||||||||||||||||||||||||||||||||||||||||
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      67 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 83
         |||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      84 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

     100 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 117
         |||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

     117 ysLysMetPhePhe 121
         |||||||||||||
     341 AGAAAATGTTCTTC 354

seq_name: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-713-067-1

seq_documentation_block:
; Sequence 1, Application US/09713067
; GENERAL INFORMATION:
;  APPLICANT:  Holloway, James L.
;  TITLE OF INVENTION:  EDUCATIONAL KIT AND METHOD USING
;  TITLE OF INVENTION:  COLIPASE-LIKE PROTEIN-1
;  FILE REFERENCE:  99-85US
;  CURRENT APPLICATION NUMBER: US/09/713,067
;  CURRENT FILING DATE:  2000-11-15
;  PRIOR APPLICATION NUMBER:  60/166,626
;  PRIOR FILING DATE:  1999-11-19
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 523
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(354)
US-09-713-067-1


alignment_scores:
            Quality:  497.00             Length:     88
              Ratio:  5.648                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179 x US-09-713-067-1   ..

 Align seg 1/1  to: US-09-713-067-1  from: 1  to: 523

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

      50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
         |||||||||||||||||||||||||||||||||||||||||||||||||
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      67 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 83
         |||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      84 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

     100 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 117
         |||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

     117 ysLysMetPhePhe 121
         |||||||||||||
     341 AGAAAATGTTCTTC 354
```



```
OM of: US-09-215-435-179 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:08 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-1210 +   77.50  153.15  0.2573
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-9 +   74.50  124.19  10.56     308
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-7 +   74.50  119.81  18.51     452
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-206 +   74.00  126.14  8.22      2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-184 +   74.00  123.12  12.12     3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-177 +   74.00  120.13  17.78     3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-353 +   74.00  117.88  23.72     4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-28 -   72.50  132.39  3.69      10
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-189-455 +   72.50  127.91  6.55      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-192-455 +   72.50  127.91  6.55      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-697-455 +   72.50  127.91  6.55      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-824-455 +   72.50  127.91  6.55      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-298-88 +   72.50  125.98  8.39      17
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-1085 +   72.50  125.98  8.39
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-344 +   72.50   95.37  425.59    2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3237 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3238 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3239 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3240 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3241 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3242 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3243 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3244 +   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3245 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3246 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3247 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3248 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3249 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3250 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3251 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3252 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3253 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3254 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3255 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3256 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3257 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3258 -   72.50   95.37  425.59
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-795-668-1381 +   71.50  140.83  1.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5020 +   71.50  140.22  1.35
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5019 +   71.50  139.35  1.51
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-2381 -   71.50  122.56  13.02
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-795-668-1 +   71.50   46.79  1.5e+05
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-6605 -   71.00  137.20  1.99
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-702 +   71.00  102.85  163.08
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-10083 -   70.00  136.95  2.06

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-1210

seq_documentation_block:
; Sequence 1210, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1210
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(437)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-1210


alignment_scores:
            Quality:   77.50              Length:    111
              Ratio:  1.490                Gaps:      8
 Percent Similarity:  46.847   Percent Identity: 31.532

alignment_block:
 US-09-215-435-179 x US-09-922-340-1210   ..

 Align seg 1/1  to: US-09-922-340-1210  from: 1  to: 437

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePhe..... 16
         :::|||      |||||||||||||||:::||||||   :::|||
     114 ATCCTGCAGGGCTGGCTGCTACTGCTATGGCTTCTACCCCTGTTCCATTA 163

      17 .............LeuPheLeuLeuThrArg...GlySerLeuSerProT 28
                      ||||||||||||                  ||||||:
     164 GCGGCAACCAGTTCTTTTTCTTCTCCAAATGTATCCTCCATGGAGTCCTT 213

      28 hrLysTyrAsn.LeuLeuGluLeuLysGluSerCysIleArgAsnGlnAs 44
         ::   :::::: :::|||
     214 CCCACCACTCGCATACTC.............................. 231

      44 pCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu..... 58
                           :::   ||||||:::::||||||
     232 .............TACTCCTCAGCCGCCCCTTCCTCCTGTGAGGCCTTCA 268

      59 ........SerHisCysAlaGluLysGlySerGluGlySerLeuCysGln 72
                :::   ||||||:::::   |||      |||:::|||
     269 GCACCATTACCTTTTGTGCCTCCTCCTGCAGT......TCCTTCTGTCCC 312

      73 ThrGlnValPhePheGlyGlnTyrArgAla.................Cy 83
         |||      :::|||      |||   :::                  ||
     313 ACCACTTGTTACTTC......TATGCCACCTCCTGTTTCTCCATCAACTG 356

      83 sProCysLeuArgAsnLeuThrCysIleTyr 93
         |   |||||||||      :::||||||||||
     357 CTGCTGCCTTCGGNAATCCTCCTGTATCTAC 387

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-9
```

```
seq_documentation_block:
; Sequence 9, Application US/09703559
; GENERAL INFORMATION:
;  APPLICANT: Emory University
;  TITLE OF INVENTION: Insulin-Responsive Sequence DNA Binding Protein-1 and Methods to
;  TITLE OF INVENTION:  Insulin-Responsive Genes
;  FILE REFERENCE: E056 1010
;  CURRENT APPLICATION NUMBER: US/09/703,559
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 3087
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1)..(3084)
US-09-703-559-9


alignment_scores:
            Quality:   74.50              Length:     63
              Ratio:  2.328                Gaps:      5
 Percent Similarity:  50.794   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179 x US-09-703-559-9   ..

 Align seg 1/1  to: US-09-703-559-9  from: 1  to: 3087

      44 AspCysGluThrGly..............CysCysGlnArgAlaProAs 55
         :::|||      |||                |||   :::|||:::|||||
     205 GAGTGCCCCGAAGGCTTCATGGGCCTGGACTGCAGGGAGAGAGTCCCCGA 254

      55 pAsnCysGlu............SerHisCysAlaGluLysGlySerGluG 68
         |:::||||||            ::::::|||         |||::::::
     255 TGACTGTGAGTGCCGCAACGGAGGCAGATGCCTG......GGCGCCAACA 298

      68 lySerLeuCysGlnThrGlnVal...PhePheGly.............. 78
         :::|||||||||      :::   |||||||||
     299 CCACCCTCTGCCAGTGCCCCCTGGGATTCTTTGGGCTTCTCTGTGAATTT 348

      79 GlnTyrArgAlaCysProCysLeuArgAsnLeuThrCys 91
         :::      |||   ||||||      |||      |||
     349 GAAATCACAGCCATGCCCTGCAACATGAACACACAGTGC 387

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-7

 seq_documentation_block:
 ; Sequence 7, Application US/09703559
 ; GENERAL INFORMATION:
 ;  APPLICANT: Emory University
 ;  TITLE OF INVENTION: Insulin-Responsive Sequence DNA Binding Protein-1 and Methods to
 ;  TITLE OF INVENTION:  Insulin-Responsive Genes
 ;  FILE REFERENCE: E056 1010
 ;  CURRENT APPLICATION NUMBER: US/09/703,559
 ;  CURRENT FILING DATE:  2001-03-12
 ;  NUMBER OF SEQ ID NOS: 43
 ;  SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 7
 ;   LENGTH: 4524
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: 5'UTR
 ;   LOCATION: (1)..(58)
 ;   NAME/KEY: exon
 ;   LOCATION: (59)..(3082)
 ;   NAME/KEY: 3'UTR
 ;   LOCATION: (3083)..(4524)
 US-09-703-559-7


alignment_scores:
            Quality:   74.50              Length:     63
              Ratio:  2.328                Gaps:      5
 Percent Similarity:  50.794   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179 x US-09-703-559-7   ..

 Align seg 1/1  to: US-09-703-559-7  from: 1  to: 4524

      44 AspCysGluThrGly..............CysCysGlnArgAlaProAs 55
         :::|||      |||                |||   :::|||:::|||||
     263 GAGTGCCCCGAAGGCTTCATGGGCCTGGACTGCAGGGAGAGAGTCCCCGA 312

      55 pAsnCysGlu............SerHisCysAlaGluLysGlySerGluG 68
         |:::||||||            ::::::|||         |||::::::
     313 TGACTGTGAGTGCCGCAACGGAGGCAGATGCCTG......GGCGCCAACA 356

      68 lySerLeuCysGlnThrGlnVal...PhePheGly.............. 78
         :::|||||||||      :::   |||||||||
     357 CCACCCTCTGCCAGTGCCCCCTGGGATTCTTTGGGCTTCTCTGTGAATTT 406

      79 GlnTyrArgAlaCysProCysLeuArgAsnLeuThrCys 91
         :::      |||   ||||||      |||      |||
     407 GAAATCACAGCCATGCCCTGCAACATGAACACACAGTGC 445

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-206

 seq_documentation_block:
 ; Sequence 206, Application US/09970966
 ; GENERAL INFORMATION:
 ;  APPLICANT: Stolk, John A.
 ;  APPLICANT:  Molesh, David Alan
 ;  APPLICANT:  Fling, Steven P.
 ;  APPLICANT:  Xu, Jiangchun
 ;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
 ;  FILE REFERENCE: 210121.484C6
 ;  CURRENT APPLICATION NUMBER: US/09/970,966
 ;  CURRENT FILING DATE:  2001-10-02
 ;  NUMBER OF SEQ ID NOS: 215
 ;  SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 206
 ;   LENGTH: 2364
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 US-09-970-966-206


 alignment_scores:
             Quality:   74.00              Length:    103
               Ratio:  1.321                Gaps:      4
  Percent Similarity:  54.369   Percent Identity: 20.388

 alignment_block:
  US-09-215-435-179 x US-09-970-966-206   ..

  Align seg 1/1  to: US-09-970-966-206  from: 1  to: 2364

       34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
          |||      |||:::|||      |||:::::::|||:::   :::   ||
     1567 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1616

       50 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 62
          |    :::                 |||:::|||::::::   |||:::
     1617 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1666

       63 ............................GluLysGlySerGluGly 68
                                      :::   ::::::|||
     1667 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1716
```

```
      69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
         |||:::|||:::::::::::::      ::::::             |||
    1717 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1766

      85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
            :::::::::   |||:::   |||      :::::::|||
    1767 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1816

     100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
           |||::::::   |||:::   ::::::      ||||||::::::
    1817 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1863

     117 LysLysMet 119
         |||::::::
    1864 AAGTCTCTG 1872

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-184

seq_documentation_block:
; Sequence 184, Application US/09970966
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184
;   LENGTH: 3079
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-184


alignment_scores:
            Quality:   74.00             Length:    103
              Ratio:  1.321               Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179 x US-09-970-966-184   ..

 Align seg 1/1  to: US-09-970-966-184  from: 1  to: 3079

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1651 GAGGAAATGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1700

      50 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 62
         |      :::                |||:::|||::::::   |||:::
    1701 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1750

      63 ..............................GluLysGlySerGluGly 68
                                                :::   ::::::|||
    1751 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1800

      69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
         |||:::|||:::::::::::::      ::::::             |||
    1801 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1850

      85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
            :::::::::   |||:::   |||      :::::::|||
    1851 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1900

     100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
           |||::::::   |||:::   ::::::      ||||||::::::
    1901 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1947

     117 LysLysMet 119
         |||::::::
    1948 AAGTCTCTG 1956

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-177

seq_documentation_block:
; Sequence 177, Application US/09970966
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 177
;   LENGTH: 3999
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-177


alignment_scores:
            Quality:   74.00             Length:    103
              Ratio:  1.321               Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179 x US-09-970-966-177   ..

 Align seg 1/1  to: US-09-970-966-177  from: 1  to: 3999

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1078 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1127

      50 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 62
         |      :::                |||:::|||::::::   |||:::
    1128 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1177

      63 ..............................GluLysGlySerGluGly 68
                                                :::   ::::::|||
    1178 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1227

      69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
         |||:::|||:::::::::::::      ::::::             |||
    1228 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1277

      85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
            :::::::::   |||:::   |||      :::::::|||
    1278 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1327

     100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
           |||::::::   |||:::   ::::::      ||||||::::::
    1328 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1374

     117 LysLysMet 119
         |||::::::
    1375 AAGTCTCTG 1383

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-353

seq_documentation_block:
; Sequence 353, Application US/09976594
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
```

```
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 353
;   LENGTH: 4866
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 415378.3
US-09-976-594-353


alignment_scores:
            Quality:   74.00              Length:    103
              Ratio:  1.321                Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179 x US-09-976-594-353   ..

 Align seg 1/1  to: US-09-976-594-353  from: 1  to: 4866

      34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
         |||      |||:::|||      |||:::::::|||:::   :::   ||
    1655 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1704

      50 sGlnArgAlaPro.............AspAsnCysGluSerHisCysAla. 62
         |      :::               |||:::|||:::::   |||:::
    1705 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1754

      63 ...............................GluLysGlySerGluGly 68
                                        :::   ::::::|||
    1755 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1804

      69 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 85
         |||:::|||:::::::::::::     ::::::          |||
    1805 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1854

      85 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 100
           :::::::::   |||:::   |||      :::::::|||
    1855 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1904

     100 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 116
           |||:::::::   |||:::   ::::::      ||||||::::::
    1905 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1951

     117 LysLysMet 119
         |||::::::
    1952 AAGTCTCTG 1960

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-28

 seq_documentation_block:
 ; Sequence 28, Application US/09886055
 ; GENERAL INFORMATION:
 ;  APPLICANT: STRYER, LUBERT
 ;  APPLICANT:  ZOZULYA, SERGEY
 ;  TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
 ;  TITLE OF INVENTION:  BIOSENSORS OF CHEMICAL SENSANTS
 ;  FILE REFERENCE: 078003-0277150
 ;  CURRENT APPLICATION NUMBER: US/09/886,055
 ;  CURRENT FILING DATE:  2001-06-22
 ;  PRIOR APPLICATION NUMBER: 60/213,812
 ;  PRIOR FILING DATE: 2000-06-22
 ;  NUMBER OF SEQ ID NOS: 522
 ;  SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ;   LENGTH: 1026
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
US-09-886-055-28


alignment_scores:
            Quality:   72.50              Length:    133
              Ratio:  1.250                Gaps:      6
 Percent Similarity:  43.609   Percent Identity: 24.812

alignment_block:
 US-09-215-435-179 x US-09-886-055-28/rev   ..

 Align seg 1/1  to reverse of: US-09-886-055-28  from: 1  to: 1026

      21 ThrArgGlySerLeuSerProThrLysTyrAsnLeuGluLeuLysGl 37
         |||   |||:::   |||||||||      :::      ::::::   ::
     798 ACCGGTGGAGGCTGCTCGCCGACTGCCCTCGACCGAGCGCATGTGGATCA 749

      37 uSer.............CysIleArgAsnGlnAspCys...GluThrGlyC 49
         :|||             :::||||||::::::|||      :::   ||||
     748 CAGCCCCAGCGATGAAGCCATAAGACACCGTGATAGCTAACACCGTGGCT 699

      49 ysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGly 65
         ||   :::::::|||   |||:::   |||:::::      |||   |||
     698 GTCTGGATGAAGCCACAGATGGCGAAGAGAAGGAGTTCATTGAGACTGGT 649

      65 ..................................................... 65

     648 GTCACTGCACGAGATGGCCAGCAGTGGAGGGATATCGCAGAAGAAGCTAT 599

      66 .....................................................S 66
                                                             :
     598 TGATCTTCCGGGAGCGGCAGAAGCTCAGGCGGAAGGTGAGGGTTGTGTGA 549

      66 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 82
         :::::|||:::   |||   ||||||:::      :::|||
     548 ACAAAGGCACTCACTGCCCCACCCAGGCCTGATGCTCCCAGCAAGGCCAG 499

      83 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 99
         :::      |||||||:::      ||||||          ||||||||
     498 GCATAGACGCTGCGACATAGCTGTTGTATA............GAGAAGTG 461

      99 pLeuSer.................IleAlaTyrGlyArgCysGlnLys 109
         |:::|||                |||::::::|||   |||||||:::
     460 GGTTTCTGATGGCCACGTAGCGGTCATAGGCCATGGC...TGCCAGCAA 415

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-189-455

 seq_documentation_block:
 ; Sequence 455, Application US/09978189
 ; GENERAL INFORMATION:
 ;  APPLICANT: Ashkenazi, Avi
 ;  APPLICANT:  Baker Kevin P.
 ;  APPLICANT:  Botstein, David
 ;  APPLICANT:  Desnoyers, Luc
 ;  APPLICANT:  Eaton, Dan
 ;  APPLICANT:  Ferrara, Napoleon
 ;  APPLICANT:  Filvaroff, Ellen
 ;  APPLICANT:  Fong, Sherman
 ;  APPLICANT:  Gao, Wei-Qiang
 ;  APPLICANT:  Gerber, Hanspeter
 ;  APPLICANT:  Gerritsen, Mary E.
 ;  APPLICANT:  Goddard, Audrey
 ;  APPLICANT:  Godowski, Paul J.
 ;  APPLICANT:  Grimaldi, J. Christopher
 ;  APPLICANT:  Gurney, Austin L.
 ;  APPLICANT:  Hillan, Kenneth J
 ;  APPLICANT:  Kljavin, Ivar J.
 ;  APPLICANT:  Kuo, Sophia S.
 ;  APPLICANT:  Napier, Mary A.
 ;  APPLICANT:  Pan, James;
 ;  APPLICANT:  Paoni, Nicholas F.
```

; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630P1C7
; CURRENT APPLICATION NUMBER: US/09/978,189
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640

```
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084598
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-5-07
;  PRIOR APPLICATION NUMBER: 60/084627
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084643
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/085339
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085338
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085323
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085582
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085700
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085689
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085580
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085573
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
            Quality:   72.50              Length:     86
              Ratio:  1.726                Gaps:      4
 Percent Similarity:  48.837   Percent Identity: 26.744

alignment_block:
 US-09-215-435-179 x US-09-978-189-455    ..

 Align seq 1/1  to: US-09-978-189-455  from: 1  to: 1518

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   :::||||||         |||:::           :::   ||
     648 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 697

      38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
         |   |||:::|||:::::||||||   :::|||   |||
     698 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT..........  737

      54 roAspAsnCysGluSerHis..............CysAlaGluLysGly 65
                |||      |||               |||
     738 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 779

      66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln........ 79
         :::||||||:::::|||      :::         |||:::
     780 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 829

      80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
                :::::   |||   |||   ::::::|||:::|||      :
     830 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 879

      94 erLysAsn 96
         ::|||:::
     880 AGAAAGAT 887

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-192-455

 seq_documentation_block:
 ; Sequence 455, Application US/09978192
 ; GENERAL INFORMATION:
 ;  APPLICANT: Ashkenazi, Avi
 ;  APPLICANT:  Baker Kevin P.
 ;  APPLICANT:  Botstein, David
 ;  APPLICANT:  Desnoyers, Luc
 ;  APPLICANT:  Eaton, Dan
 ;  APPLICANT:  Ferrara, Napoleon
 ;  APPLICANT:  Filvaroff, Ellen
 ;  APPLICANT:  Fong, Sherman
 ;  APPLICANT:  Gao, Wei-Qiang
 ;  APPLICANT:  Gerber, Hanspeter
 ;  APPLICANT:  Gerritsen, Mary E.
 ;  APPLICANT:  Goddard, Audrey
 ;  APPLICANT:  Godowski, Paul J.
 ;  APPLICANT:  Grimaldi, J. Christopher
 ;  APPLICANT:  Gurney, Austin L.
 ;  APPLICANT:  Hillan, Kenneth J
 ;  APPLICANT:  Kljavin, Ivar J.
 ;  APPLICANT:  Kuo, Sophia S.
 ;  APPLICANT:  Napier, Mary A.
 ;  APPLICANT:  Pan, James;
 ;  APPLICANT:  Paoni, Nicholas F.
 ;  APPLICANT:  Roy, Margaret Ann
 ;  APPLICANT:  Shelton, David L.
 ;  APPLICANT:  Stewart, Timothy A.
 ;  APPLICANT:  Tumas, Daniel
 ;  APPLICANT:  Williams, P. Mickey
 ;  APPLICANT:  Wood, William I.
 ;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ;  TITLE OF INVENTION:  Acids Encoding the Same
 ;  FILE REFERENCE: P2630P1C9
 ;  CURRENT APPLICATION NUMBER: US/09/978,192
 ;  CURRENT FILING DATE:  2001-10-15
 ;  PRIOR APPLICATION NUMBER: 09/918585
 ;  PRIOR FILING DATE: 2001-07-30
 ;  PRIOR APPLICATION NUMBER: 60/062250
 ;  PRIOR FILING DATE: 1997-10-17
 ;  PRIOR APPLICATION NUMBER: 60/064249
 ;  PRIOR FILING DATE: 1997-11-03
 ;  PRIOR APPLICATION NUMBER: 60/065311
 ;  PRIOR FILING DATE: 1997-11-13
 ;  PRIOR APPLICATION NUMBER: 60/066364
 ;  PRIOR FILING DATE: 1997-11-21
 ;  PRIOR APPLICATION NUMBER: 60/077450
 ;  PRIOR FILING DATE: 1998-03-10
 ;  PRIOR APPLICATION NUMBER: 60/077632
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077641
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077649
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077791
 ;  PRIOR FILING DATE: 1998-03-12
 ;  PRIOR APPLICATION NUMBER: 60/078004
 ;  PRIOR FILING DATE: 1998-03-13
 ;  PRIOR APPLICATION NUMBER: 60/078886
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078936
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078910
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078939
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/079294
 ;  PRIOR FILING DATE: 1998-03-25
 ;  PRIOR APPLICATION NUMBER: 60/079656
 ;  PRIOR FILING DATE: 1998-03-26
 ;  PRIOR APPLICATION NUMBER: 60/079664
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079689
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079663
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079728
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079786
```

; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-5-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15

alignment_scores:
Quality: 72.50 Length: 86
Ratio: 1.726 Gaps: 4
Percent Similarity: 48.837 Percent Identity: 26.744

alignment_block:
US-09-215-435-179 x US-09-978-192-455 ..

Align seg 1/1 to: US-09-978-192-455 from: 1 to: 1518

```
  22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
     |||   :::||||||         |||:::            :::   ||
 648 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 697

  38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
     |   |||:::|||:::::::||||||   :::|||   |||
 698 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.......... 737

  54 roAspAsnCysGluSerHis..............CysAlaGluLysGly 65
        |||      |||               |||
 738 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 779
```

```
     66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln..... 79
        :::|||||||::::::|||      :::         |||:::
    780 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 829

     80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
               ::::::   |||   |||   ::::::|||:::|||     :
    830 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 879

     94 erLysAsn 96
        ::|||:::
    880 AGAAAGAT 887

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-697-455

seq_documentation_block:
; Sequence 455, Application US/09978697
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi, Avi
;  APPLICANT:  Baker Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan
;  APPLICANT:  Ferrara, Napoleon
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, J. Christopher
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Kuo, Sophia S.
;  APPLICANT:  Napier, Mary A.
;  APPLICANT:  Pan, James;
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Shelton, David L.
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2630P1C27
;  CURRENT APPLICATION NUMBER: US/09/978,697
;  CURRENT FILING DATE:  2001-10-16
;  PRIOR APPLICATION NUMBER: 09/918585
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064249
;  PRIOR FILING DATE: 1997-11-03
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066364
;  PRIOR FILING DATE: 1997-11-21
;  PRIOR APPLICATION NUMBER: 60/077450
;  PRIOR FILING DATE: 1998-03-10
;  PRIOR APPLICATION NUMBER: 60/077632
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077641
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077649
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077791
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/078004
;  PRIOR FILING DATE: 1998-03-13
;  PRIOR APPLICATION NUMBER: 60/078886
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078936
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078939
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/079294
;  PRIOR FILING DATE: 1998-03-25
;  PRIOR APPLICATION NUMBER: 60/079656
;  PRIOR FILING DATE: 1998-03-26
;  PRIOR APPLICATION NUMBER: 60/079664
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079663
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079728
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079786
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079920
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/079923
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/080105
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080107
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080165
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080194
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080327
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080328
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080333
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080334
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/081070
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081049
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081071
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081195
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081203
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081229
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081955
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081817
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081819
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081952
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081838
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/082568
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/082569
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/082704
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082804
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082700
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082797
```

```
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082796
;  PRIOR FILING DATE: 1998-04-23
;  PRIOR APPLICATION NUMBER: 60/083336
;  PRIOR FILING DATE: 1998-04-27
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/083392
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083495
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083496
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083499
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083545
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083554
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083558
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083559
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083500
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083742
;  PRIOR FILING DATE: 1998-04-30
;  PRIOR APPLICATION NUMBER: 60/084366
;  PRIOR FILING DATE: 1998-05-05
;  PRIOR APPLICATION NUMBER: 60/084414
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: 60/084441
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: 60/084637
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084639
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084640
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084598
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-5-07
;  PRIOR APPLICATION NUMBER: 60/084627
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084643
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/085339
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085338
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085323
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085582
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085700
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085689
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085580
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085573
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
            Quality:   72.50          Length:     86
              Ratio:  1.726            Gaps:      4
 Percent Similarity:  48.837   Percent Identity: 26.744

alignment_block:
 US-09-215-435-179 x US-09-978-697-455   ..

 Align seg 1/1  to: US-09-978-697-455  from: 1  to: 1518

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   :::|||||||         |||:::             :::   ||
     648 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 697

      38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
         |   |||:::|||:::::::||||||   :::|||   |||
     698 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.......... 737

      54 roAspAsnCysGluSerHis..............CysAlaGluLysGly 65
                |||      |||               |||
     738 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 779

      66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln........ 79
         :::||||||:::::::|||      :::          |||:::
     780 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 829

      80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
                 ::::::   |||   |||   ::::::|||:::|||      :
     830 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 879

      94 erLysAsn 96
         ::|||:::
     880 AGAAAGAT 887

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-824-455

seq_documentation_block:
; Sequence 455, Application US/09978824
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi, Avi
;  APPLICANT:  Baker Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan
;  APPLICANT:  Ferrara, Napoleon
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, J. Christopher
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Kuo, Sophia S.
;  APPLICANT:  Napier, Mary A.
;  APPLICANT:  Pan, James;
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Shelton, David L.
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2630P1C14
;  CURRENT APPLICATION NUMBER: US/09/978,824
;  CURRENT FILING DATE:  2001-10-17
;  PRIOR APPLICATION NUMBER: 09/918585
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064249
```

; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-5-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582

```
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085700
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085689
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085580
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085573
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
            Quality:   72.50             Length:     86
              Ratio:  1.726               Gaps:      4
 Percent Similarity:  48.837   Percent Identity: 26.744

alignment_block:
 US-09-215-435-179 x US-09-978-824-455   ..

 Align seg 1/1  to: US-09-978-824-455  from: 1  to: 1518

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   :::||||||         |||:::            :::   ||
     648 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 697

      38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
         |   |||:::|||::::::||||||   :::|||   |||
     698 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.......... 737

      54 roAspAsnCysGluSerHis..............CysAlaGluLysGly 65
                |||      |||               |||
     738 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 779

      66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln........ 79
         :::||||||::::::|||      :::         |||:::
     780 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 829

      80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
                :::::   |||   |||   ::::::|||:::|||      :
     830 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 879

      94 erLysAsn 96
         ::|||:::
     880 AGAAAGAT 887

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-298-88

seq_documentation_block:
; Sequence 88, Application US/09974298
; GENERAL INFORMATION:
;  APPLICANT: Chen, Huei-Mei
;  TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
;  FILE REFERENCE: PA-0037 P
;  CURRENT APPLICATION NUMBER: US/09/974,298
;  CURRENT FILING DATE:  2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,331
;  PRIOR FILING DATE: 2000-05-10
;  NUMBER OF SEQ ID NOS: 194
;  SOFTWARE: PERL Program
; SEQ ID NO 88
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 2481150CB1
US-09-974-298-88


alignment_scores:
            Quality:   72.50             Length:     86
              Ratio:  1.726               Gaps:      4
 Percent Similarity:  48.837   Percent Identity: 26.744

alignment_block:
 US-09-215-435-179 x US-09-974-298-88   ..

 Align seg 1/1  to: US-09-974-298-88  from: 1  to: 1797

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   :::||||||         |||:::            :::   ||
     662 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 711

      38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
         |   |||:::|||::::::||||||   :::|||   |||
     712 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.......... 751

      54 roAspAsnCysGluSerHis..............CysAlaGluLysGly 65
                |||      |||               |||
     752 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 793

      66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln........ 79
         :::||||||::::::|||      :::         |||:::
     794 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 843

      80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
                :::::   |||   |||   ::::::|||:::|||      :
     844 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 893

      94 erLysAsn 96
         ::|||:::
     894 AGAAAGAT 901

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-1085

seq_documentation_block:
; Sequence 1085, Application US/09976594
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 1085
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 2481150CB1
US-09-976-594-1085


alignment_scores:
            Quality:   72.50             Length:     86
              Ratio:  1.726               Gaps:      4
 Percent Similarity:  48.837   Percent Identity: 26.744

alignment_block:
 US-09-215-435-179 x US-09-976-594-1085   ..

 Align seg 1/1  to: US-09-976-594-1085  from: 1  to: 1797

      22 ArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSe 38
         |||   :::||||||         |||:::            :::   ||
     662 AGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTC 711
```

```
     38 r...CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaP 54
        |   |||:::|||:::::::|||||   :::|||   |||
    712 TGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT..... 751

     54 roAspAsnCysGluSerHis.............CysAlaGluLysGly 65
           |||      |||              |||
    752 ........TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTG 793

     66 SerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln........ 79
        :::|||||||:::::::|||      :::         |||:::
    794 AAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACT 843

     80 .......TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrS 94
              ::::::   |||   |||   ::::::|||:::|||      :
    844 AGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATAC 893

     94 erLysAsn 96
        ::|||:::
    894 AGAAAGAT 901
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-344

```
seq_documentation_block:
; Sequence 344, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 344
;   LENGTH: 26092
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12281)..(13480)
US-09-897-516-344


alignment_scores:
            Quality:   72.50            Length:    142
              Ratio:  1.098              Gaps:      9
 Percent Similarity:  46.479   Percent Identity: 26.056

alignment_block:
 US-09-215-435-179 x US-09-897-516-344   ..

 Align seg 1/1  to: US-09-897-516-344  from: 1  to: 26092

      5 GlnTrpLeuLeuLeu.............................LeuPh 11
        |||||||||||||||                              :::::
  19709 CAGTGGTTACTCCTTACCCGCCACAGATAACTGCTGGTAAATATAATGCT 19758

     11 eLeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProT 28
        ::::   |||:::      ::::::||||||      :::|||   ||||
  19759 TATAGCTTTCTACCAGCGCCTTATCCTGACAGTCGTTAAGCTCCCCCCCA 19808

     28 hrLysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAsp 44
        ||:::        :::        :::   |||:::
  19809 CACAA.........ATGACGCATGGTCATATTTGCCTG............ 19837

     45 CysGluThrGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCy 61
           |||:::|||   :::|||      |||   :::   :::::::|||
  19838 ATAGAAAGCGGGAACGCTCAA......CCGTGCTCTTTCCAAAATCATAC 19881

     61 sAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnValPhePheG 78
        :::||||||   :::   |||:::                    |||   |
  19882 CACCGAGAAAAGAGATATCGGTCAG.................TTTACCG 19913

     78 lyGlnTyrArg..................................... 81
        ||:::||||||
  19914 GTCGTTATCGGCTCTCCCTGTTTCAGGTTATCCCTCATGGTAAATTGCGT 19963

     82 ...AlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGl 97
              |||      :::|||:::         :::      |||   ::
  19964 CAGATGTGCCGG...ATCAGAAAGGGTAGAAGCGTAATCGTTAAAATGGA 20010

     97 uLysTrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArg.... 112
        :   ||||||   |||:::      :::|||:::|||   ||||||
  20011 TATTTGGCTGATGATCACC......GAATGTGAGAAAAATGGTCGGTTTG 20054

    113 ...........GlnLysLeuAlaLys 117
                   :::||||||:::|||
  20055 TCGCGTTTAGCCACAAACTCACTAAA 20080
```

```
OM of: US-09-215-435-179 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:30 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179
Query length: 121
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est15:AI018488      -  491.50   951.42   7.1e-44   491  ! AI018488 ov63f08.s1 Soares_test
gb_gss29:AZ575911      +  196.00   386.93   2.0e-12   127  ! AZ575911 AST-T22D0047 Genetrap
gb_est88:BF509739      -  164.50   316.41   1.7e-08   293  ! BF509739 UI-H-BI4-apg-b-05-0-UI
gb_est69:BE047055      -  143.50   268.83   7.4e-06   540  ! BE047055 hq55g09.x1 NCI_CGAP_Pa
gb_est69:BE047063      -  137.00   256.31   3.7e-05   529  ! BE047063 hq55h07.x1 NCI_CGAP_Pa
gb_est81:BE969660      +  136.50   255.35   4.2e-05   528  ! BE969660 601679794F1 NIH_MGC_78
gb_est81:BE969913      +  135.50   253.29   5.5e-05   533  ! BE969913 601680001F1 NIH_MGC_78
gb_est81:BE969784      +  135.50   253.10   5.6e-05   543  ! BE969784 601679544F1 NIH_MGC_78
gb_est81:BE969920      +  135.50   252.46   6.1e-05   577  ! BE969920 601680009F1 NIH_MGC_78
gb_est81:BE969726      +  135.50   250.69   7.6e-05   682  ! BE969726 601679676F1 NIH_MGC_78
gb_est81:BE969942      +  135.50   250.13   8.2e-05   719  ! BE969942 601680032F1 NIH_MGC_78
gb_est81:BE970098      +  135.50   249.93   8.4e-05   733  ! BE970098 601679841F1 NIH_MGC_78
gb_est81:BE970031      +  135.50   249.66   8.7e-05   752  ! BE970031 601680144F1 NIH_MGC_78
gb_est81:BE970121      +  135.50   249.49   8.9e-05   764  ! BE970121 601679866F1 NIH_MGC_78
gb_est81:BE970243      +  135.50   249.38   9.0e-05   772  ! BE970243 601679987F1 NIH_MGC_78
gb_est19:AI366744      -  135.00   253.33   5.4e-05   484  ! AI366744 qq41a03.x1 Soares_NhHM
gb_est88:BF507781      -  135.00   253.01   5.7e-05   499  ! BF507781 UI-H-BI4-apv-b-04-0-UI
gb_est49:AW583164      +  131.00   247.20   0.0001    412  ! AW583164 ia11f06.y1 Human Pancr
gb_est97:BG141967      -  128.50   241.80   0.0002    432  ! BG141967 ia85b12.x1 Melton Mous
gb_est49:AW584008      -  127.50   237.91   0.0004    519  ! AW584008 ia08f03.x1 Human Pancr
gb_est81:BE970103      +  127.50   234.13   0.0006    742  ! BE970103 601679846F1 NIH_MGC_78
gb_htc:AK019005        +  125.50   234.51   0.0006    494  ! AK019005 Mus musculus 10 day ol
gb_est19:AI411581      -  125.00   237.08   0.0004    353  ! AI411581 EST239875 Normalized r
gb_est53:AW966719      +  124.50   232.88   0.0007    479  ! AW966719 EST378793 MAGE reseque
gb_est81:BE970191      +  124.50   231.50   0.0009    546  ! BE970191 601679930F1 NIH_MGC_78
gb_est68:BB565103      +  124.00   237.23   0.0004    289  ! BB565103 BB565103 RIKEN full-le
gb_est49:AW583194      +  123.50   230.03   0.0011    521  ! AW583194 ia12c01.y1 Human Pancr
gb_est81:BE970198      +  123.50   229.35   0.0012    556  ! BE970198 601679939F1 NIH_MGC_78
gb_est99:BG311607      +  123.00   229.51   0.0012    499  ! BG311607 ib26f02.y1 Melton Mous
gb_est19:AI385475      +  122.50   230.30   0.0010    422  ! AI385475 mj69b11.y1 Soares mous
gb_est98:BG270511      +  122.50   230.10   0.0011    430  ! BG270511 ia85b12.y1 Melton Mous
gb_est18:AI326929      -  122.50   229.81   0.0011    442  ! AI326929 mj69b11.x1 Soares mous
gb_est97:BG141777      -  122.50   229.79   0.0011    443  ! BG141777 ia94b06.x1 Melton Mous
gb_est99:BG311902      -  122.50   229.57   0.0011    452  ! BG311902 ia98a11.x1 Melton Mous
gb_est99:BG312118      -  122.50   228.96   0.0012    479  ! BG312118 ib03c06.x1 Melton Mous
gb_est99:BG311196      +  122.50   228.83   0.0013    485  ! BG311196 ib21c03.y1 Melton Mous
gb_est99:BG311655      +  122.50   228.83   0.0013    485  ! BG311655 ib27d10.y1 Melton Mous
gb_est98:BG270331      +  122.50   228.76   0.0013    488  ! BG270331 ib13e11.y1 Melton Mous
gb_est98:BG270318      +  122.50   228.74   0.0013    489  ! BG270318 ib13c10.y1 Melton Mous
gb_htc:AK019047        +  122.50   228.70   0.0013    491  ! AK019047 Mus musculus adult mal
gb_htc:AK008874        +  122.50   228.68   0.0013    492  ! AK008874 Mus musculus adult mal
gb_htc:AK008879        +  122.50   228.68   0.0013    492  ! AK008879 Mus musculus adult
gb_htc:AK008635        +  122.50   228.66   0.0013    493  ! AK008635 Mus musculus adult
gb_htc:AK008834        +  122.50   228.66   0.0013    493  ! AK008834 Mus musculus adult
gb_htc:AK008839        +  122.50   228.66   0.0013    493  ! AK008839 Mus musculus adult

seq_name: gb_est15:AI018488

seq_documentation_block:
LOCUS       AI018488      491 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov63f08.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642023
            3', mRNA sequence.
ACCESSION   AI018488
VERSION     AI018488.1  GI:3233007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 630   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1642023"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      129 a    115 c    133 g    113 t      1 others
ORIGIN


alignment_scores:
            Quality:  491.50            Length:    108
              Ratio:  4.819               Gaps:      3
 Percent Similarity:  94.444   Percent Identity: 91.667

alignment_block:
 US-09-215-435-179 x AI018488/rev   ..

 Align seg 1/1  to reverse of: AI018488  from: 1  to: 491

      16 PheLeuPheLeuLeuThrArgGly.SerLeuSerProThrLysTyrAsnL 32
         |||||||||||||||||||||:::  :::   ||||||||||||||||||
     470 TTTCTCTTCCTCCTCACCAGGGCTCACT...TCTCCAACAAAATACAACC 424

      32 euLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGly 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    423 TTTTGGAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGC 374

     49 CysCysGlnArgAlaProAspAsnCysGluSerHis.CysAlaGluLysG 65
        |||||||||||||||||||||||||||||||||||||    ||||||||||
    373 TGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTNGCGCGGAGAAGG 324

     65 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArg 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    323 GGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGA 274

     82 AlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLy 98
        ||||||||||||   |||||||||||||||||||||||||||||||||||
    273 GCGTGTCCCTGCTG.CGGAACCTGACTTGTATATATTCAAAGAATGAGAA 225

     98 sTrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysL 115
        ||||||||||:::   |||||||||||||||||||||||||||||||||||
    224 ATGGCTTAGCATGCCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGT 175

     115 euAlaLysLysMetPhePhe 121
         |||||||||||||||||||
    174 TGGCTAAGAAAATGTTCTTC 155
```

seq_name: gb_gss29:AZ575911

seq_documentation_block:

```
LOCUS       AZ575911      127 bp    DNA             GSS       06-DEC-2000
DEFINITION  AST-T22D0047 Genetrap T47D Human Breast Carcinoma Library Homo
            sapiens genomic 5', DNA sequence.
ACCESSION   AZ575911
VERSION     AZ575911.1  GI:11562222
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127)
  AUTHORS   Henkel,G., Liyanage,M., Pratt,E., Huang,D., Riley,M., Bernardino,A.
            , Durick,K. and Pollok,B.
  TITLE     Exon-trap tags from a T47D GenomeScreen(TM) Library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Greg Henkel
            Gene Expression
            Aurora Biosciences Corp.
            11010 Torreyana Road, San Diego, CA 92121, USA
            Tel: 8584048436
            Fax: 8584046719
            Email: henkelg@aurorabio.com
            Pools of cells were isolated from a GenomeScreen(TM) library. The
            library of cells was generated by retroviral integration of a gene
            tagging element consisting of: 1) A promoterless beta-lactamase
            proceeded by a splice acceptor as a reporter for gene expression;
            2) A promoter driving neomycin resistance followed by a splice
            donor to trap downstream exons. 3' RACE from neomycin gene was
            performed using total RNA from isolated pools. Output was shotgun
            cloned in pAmp-1 and used to transform DH5-alpha competent
            bacteria. 5' ends of reported sequences were immediately preceded
            by splice donor from the trapping construct.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .127
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Genetrap T47D Human Breast Carcinoma Library"
                     /tissue_type="Carcinoma"
                     /cell_type="Epithelial"
                     /cell_line="T47D"
                     /note="Organ: Breast; Vector: pAmp-1; 3' RACE of total RNA
                     from genetrap pools; shotgun clone in pAmp-1 and  used to
                     transform DH5-alpha competent bacteria."
BASE COUNT       28 a     34 c     43 g     21 t      1 others
ORIGIN
```

alignment_scores:

```
            Quality:  196.00            Length:     40
              Ratio:  5.026               Gaps:      0
 Percent Similarity:  97.500   Percent Identity: 97.500
```

alignment_block:

```
US-09-215-435-179 x AZ575911   ..

Align seg 1/1  to: AZ575911  from: 1  to: 127

     34 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 50

     50 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 67
        ||||||||||||||||   |||||||||||||||||||||||||||||||
     51 CCAACGTGCTCCAGACAA.TGCGAGTCGCACTGCGCGGAGAAGGGGTCCN 99

     67 luGlySerLeuCysGlnThr 73
        |||||||||||||||||
    100 AGGGCAGTCTGTGTCAAACG 119
```

seq_name: gb_est88:BF509739

seq_documentation_block:

```
LOCUS       BF509739      293 bp    mRNA            EST       06-DEC-2000
DEFINITION  UI-H-BI4-apg-b-05-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3087296 3', mRNA sequence.
ACCESSION   BF509739
VERSION     BF509739.1  GI:11593037
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3087296"
                     /clone_lib="NCI_CGAP_Sub8"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive  hybridization with a driver
                     comprising: a pool of clones from  NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2732833-2737415, 3068040-3069191; 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729326; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728969-2733190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238,  3081864-3084550
                     ; 25% of the driver population).  Subtraction was
                     performed as previously described [Bonaldo, Lennon &
```

```
                    Soares (1996): Normalization and Subtraction: Two
                    Approaches To Facilitate Gene Discovery. Genome Research
                    6, 791-806.
                    TAG_LIB=NCI_CGAP_Br2
                    TAG_TISSUE=breast
                    TAG_SEQ=AAACC"
BASE COUNT       63 a     75 c     72 g     83 t
ORIGIN


alignment_scores:
           Quality:  164.50             Length:     38
             Ratio:  5.306                Gaps:      1
 Percent Similarity:  81.579   Percent Identity: 78.947

alignment_block:
 US-09-215-435-179 x BF509739/rev   ..

 Align seg 1/1  to reverse of: BF509739  from: 1  to: 293

      48 GlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLy 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 GGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAA 235

      64 sGlySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrA 81
         ||||||||||||||||||||||||||||||:::
     234 GGGGTCCGAGGGCAGTCTGTGTCAAACGCAGCTG.............C 200

      81 rgAlaCysProCys 85
           |||   ||||||
     199 CAGCTACTCCATGC 186

seq_name: gb_est69:BE047055

seq_documentation_block:
LOCUS       BE047055      540 bp    mRNA            EST       08-JUN-2000
DEFINITION  hq55g09.x1 NCI_CGAP_Pan3 Homo sapiens cDNA clone IMAGE:3123328 3'
            similar to gb:J02883 COLIPASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   BE047055
VERSION     BE047055.1  GI:8364108
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3123328"
                     /clone_lib="NCI_CGAP_Pan3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      123 a    131 c    161 g    124 t      1 others
ORIGIN


alignment_scores:
           Quality:  143.50             Length:    113
             Ratio:  2.080                Gaps:      3
 Percent Similarity:  61.062   Percent Identity: 30.973

alignment_block:
 US-09-215-435-179 x BE047055/rev   ..

 Align seg 1/1  to reverse of: BE047055  from: 1  to: 540

       7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGl 23
         |||||||||:::   |||:::      :::      |||      |||||
     500 CTCCTGCTTGTCGCCCTTTTTGTGGCCTATGAGGCTCTGGCCCCCCGGGG 451

      23 ySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysI 40
         |   :::                    ||||||     |||   |||:
     450 GATCATTATCAAC.............TTGGAGACCGGTGAGCTCTGCA 416

      40 leArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsn 56
         ::::::::   :::|||:::::::::|||||||||:::::
     415 TGAATAGTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGTTG 366

      57 CysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnTh 73
            :::::::|||:::::::   :::|||||||:::|||   |||:::::
     365 GGCTTGGCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGT 316

      73 rGlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuT 90
         :::::::::::::|||   |||      |||||||||   |||:::||||
     315 CAAGACGCTTTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGA 266

      90 hrCysIleTyrSerLysAsnGluLysTrpLeu.............Ser 101
         |||||         ::::::::::|||   :::              :::
     265 CCTGT.........GAGGGAGACAAGACCATTGTGGGCTCCATCACCAAC 225

     102 IleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLys 114
            ::::|||   |||:::      ||||||:::|||
     224 ACCAACTTTGGCATTTGCCATGACGNTGGACGCTCCAAG 186

seq_name: gb_est69:BE047063

seq_documentation_block:
LOCUS       BE047063      529 bp    mRNA            EST       08-JUN-2000
DEFINITION  hq55h07.x1 NCI_CGAP_Pan3 Homo sapiens cDNA clone IMAGE:3123325 3'
            similar to gb:J02883 COLIPASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   BE047063
VERSION     BE047063.1  GI:8364116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 412.
```

```
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3123325"
                     /clone_lib="NCI_CGAP_Pan3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      117 a    128 c    167 g    117 t
ORIGIN


alignment_scores:
            Quality:  137.00            Length:    116
              Ratio:  1.957               Gaps:      4
 Percent Similarity:  60.345   Percent Identity: 30.172

alignment_block:
 US-09-215-435-179 x BE047063/rev   ..

 Align seg 1/1  to reverse of: BE047063  from: 1  to: 529

       9 LeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu..............  20
         :::|||:::|||||||:::      |||||||:::
     499 ATCTTGATCCTCCTGCTTGTCGCCCTCTTTGTGGCCTATGCAGCTCCTGG 450

      21 .ThrArgGlySerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysG  37
          ||||||   :::                              ||||||      |
     449 CCCCCGGGGGATCATTATCAAC..............CTGGAGAACGGTG 415

      37 luSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAla  53
         ||   |||:::::::::   :::|||:::::::::|||||||||:::::
     414 AGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCA 365

      54 ProAspAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLe  70
                    ::::::|||::::::   :::|||||||:::|||
     364 AGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCAGCGAGAACAGCGA 315

      70 uCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLeuA  87
          |||:::::::::::::::::::|||   |||      |||||||||   |
     314 GTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAGTGTCCCTGTGAGC 265

      87 rgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeu..........  100
         ||:::||||||||||         :::::::::|||   :::
     264 GTGGCCTGACCTGT.........GAGGGAGACAAGACCATCGTGGGCTCC 224

     101 ......SerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLys  114
               :::      :::|||   |||:::      ||||||:::|||
     223 ATCACCAACACCAACTTTGGCATCTGCCATGACGCTGGACGCTCCAAG 176

 seq_name: gb_est81:BE969660

 seq_documentation_block:
 LOCUS       BE969660      528 bp    mRNA            EST       04-OCT-2000
 DEFINITION  601679794F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949990 5',
             mRNA sequence.
 ACCESSION   BE969660
 VERSION     BE969660.1  GI:10582593
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 528)
   AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
   TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
   JOURNAL   Unpublished (1999)
 COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM815  row: 1  column: 23
             High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949990"
                     /clone_lib="NIH_MGC_78".
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      117 a    168 c    133 g    110 t
ORIGIN


alignment_scores:
            Quality:  136.50            Length:    111
              Ratio:  1.923               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969660   ..

 Align seg 1/1  to: BE969660  from: 1  to: 528

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh  28
         :::|||:::::::::|||      |||:::   :::      :::|||
      26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCCATGCAGCTCCTGG  75

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA  42
                ::::::          ||||||     |||   |||:::::::
      76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu  58
         ::   :::|||:::::::::|||||||||:::::
     126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa  75
         ::::::|||::::::   :::||||||:::|||   |||:::::::::::
     176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI  92
         :::::::|||   |||      |||||||||   |||:::|||||||||
     226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
                ::::::::::|||   :::                 :::
     275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      ||||||:::|||
     317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

 seq_name: gb_est81:BE969913

 seq_documentation_block:
 LOCUS       BE969913      533 bp    mRNA            EST       04-OCT-2000
 DEFINITION  601680001F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950113 5',
             mRNA sequence.
```

```
ACCESSION   BE969913
VERSION     BE969913.1  GI:10582846
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: b  column: 02
            High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950113"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      118 a    169 c    132 g    114 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969913   ..

 Align seg 1/1  to: BE969913  from: 1  to: 533

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
                :::::          ||||||      |||   |||:::::::
      76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::::
     126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         :::::::|||:::::::   :::||||||:::|||   |||:::::::::::
     176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         ::::::::|||   |||      |||||||||   |||:::|||||||||
     226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

      92 leTyrSerLysAsnGluLysTrpLeu.............SerIleAla 103
                 :::::::::|||   :::               :::
     275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      ||||||:::|||
     317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE969784

seq_documentation_block:
LOCUS       BE969784      543 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679544F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949859 5',
            mRNA sequence.
ACCESSION   BE969784
VERSION     BE969784.1  GI:10582717
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM815  row: g  column: 12
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949859"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      118 a    176 c    133 g    115 t      1 others
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969784   ..

 Align seg 1/1  to: BE969784  from: 1  to: 543

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
                :::::          ||||||      |||   |||:::::::
```

```
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
        ::   :::|||::::::::::|||||||||:::::::
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
        :::::::|||::::::   :::||||||:::|||   |||:::::::::::
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
        ::::::::|||   |||      |||||||||   |||:::|||||||||
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
                ::::::::::|||   :::                :::
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
        :::|||   |||:::      ||||||:::|||
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349
```

seq_name: gb_est81:BE969920

seq_documentation_block:

```
LOCUS       BE969920      577 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680009F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950115 5',
            mRNA sequence.
ACCESSION   BE969920
VERSION     BE969920.1  GI:10582853
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: b  column: 04
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950115"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      139 a    174 c    144 g    120 t
ORIGIN
```

```
alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908                 Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969920   ..

 Align seg 1/1  to: BE969920  from: 1  to: 577

     12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
        :::|||:::::::::|||      |||:::   :::      :::|||
     26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

     28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
                :::::          ||||||      |||   |||::::::
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
        ::   :::|||::::::::::|||||||||:::::::
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
        :::::::|||::::::   :::||||||:::|||   |||:::::::::::
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
        ::::::::|||   |||      |||||||||   |||:::|||||||||
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu.....................SerIleAla 103
                ::::::::::|||   :::                   :::
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
        :::|||   |||:::      ||||||:::|||
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349
```

seq_name: gb_est81:BE969726

seq_documentation_block:

```
LOCUS       BE969726      682 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679676F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949891 5',
            mRNA sequence.
ACCESSION   BE969726
VERSION     BE969726.1  GI:10582659
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM815  row: h  column: 20
            High quality sequence stop: 547.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949891"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
```

```
                    ); 5' and 3' adaptors were used in cloning as follows: 5'
                    adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                    sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                    (where B = A, C, or G and N = A, C,  G, or T).   Average
                    insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                    contained inserts  by PCR. This library was enriched for
                    full-length clones and was constructed by Clontech
                    Laboratories (Palo Alto, CA)."
BASE COUNT      173 a    198 c    191 g    120 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969726   ..

 Align seg 1/1  to: BE969726  from: 1  to: 682

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
               ::::::          ||||||      |||   |||:::::::
      76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::::
     126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         :::::::|||:::::::   :::||||||:::|||   |||:::::::::::
     176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         ::::::::|||   |||      |||||||||   |||:::|||||||||
     226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
               ::::::::::|||   :::                :::
     275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      |||||||:::|||
     317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE969942

seq_documentation_block:
LOCUS       BE969942      719 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680032F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950455 5',
            mRNA sequence.
ACCESSION   BE969942
VERSION     BE969942.1  GI:10582875
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: p  column: 08
            High quality sequence stop: 524.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950455" .
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      160 a    233 c    191 g    135 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE969942   ..

 Align seg 1/1  to: BE969942  from: 1  to: 719

      12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
         :::|||:::::::::|||      |||:::   :::      :::|||
      26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

      28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
               ::::::          ||||||      |||   |||:::::::
      76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

      42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
         ::   :::|||:::::::::|||||||||:::::::
     126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

      59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
         :::::::|||:::::::   :::||||||:::|||   |||:::::::::::
     176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

      75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
         :::::::|||   |||      |||||||||   |||:::|||||||||
     226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

      92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
               ::::::::::|||   :::                :::
     275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

     104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
         :::|||   |||:::      |||||||:::|||
     317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE970098

seq_documentation_block:
LOCUS       BE970098      733 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679841F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950098 5',
            mRNA sequence.
ACCESSION   BE970098
VERSION     BE970098.1  GI:10583131
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: a  column: 11
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950098"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      171 a    218 c    210 g    134 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE970098   ..

 Align seg 1/1  to: BE970098  from: 1  to: 733

     12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
        :::|||:::::::::|||      |||:::   :::      :::|||
     26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

     28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
              ::::::         ||||||      |||   |||:::::::
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
        ::   :::|||:::::::::|||||||||:::::
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
        :::::::|||::::::   :::|||||||:::|||   |||:::::::::::
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
        :::::::|||   |||      |||||||||   |||:::||||||||||
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
              ::::::::::|||   :::               :::
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
        :::|||   |||:::      ||||||:::|||
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE970031

seq_documentation_block:
LOCUS       BE970031      752 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680144F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950266 5',
            mRNA sequence.
ACCESSION   BE970031
VERSION     BE970031.1  GI:10582964
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: h  column: 11
            High quality sequence stop: 523.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950266"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      168 a    255 c    201 g    128 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE970031   ..

 Align seg 1/1  to: BE970031  from: 1  to: 752

     12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
        :::|||:::::::::|||      |||:::   :::      :::|||
     26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

     28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
              ::::::         ||||||      |||   |||:::::::
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
```

```
      ::   :::|||::::::::::|||||||||::::::
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
        :::::|||::::::   :::|||||||:::|||   |||::::::::::
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
        :::::::|||   |||      |||||||||   |||:::|||||||||
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu.............SerIleAla 103
                :::::::::|||   :::               :::
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
        :::|||   |||:::      ||||||:::||| 
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE970121

seq_documentation_block:
LOCUS       BE970121      764 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679866F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950152 5',
            mRNA sequence.
ACCESSION   BE970121
VERSION     BE970121.1  GI:10583154
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: c  column: 17
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950152"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      175 a    241 c    211 g    137 t
ORIGIN


alignment_scores:
            Quality:  135.50            Length:    111
              Ratio:  1.908               Gaps:      3
 Percent Similarity:  63.964   Percent Identity: 28.829

alignment_block:
 US-09-215-435-179 x BE970121   ..

 Align seg 1/1  to: BE970121  from: 1  to: 764

     12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
        :::|||::::::::::|||      |||:::  :::      :::|||
     26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

     28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
              ::::::         ||||||      |||   |||:::::::
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
        ::   :::|||::::::::::|||||||||::::::
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
        ::::::|||::::::   :::||||||:::|||   |||::::::::::
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
        :::::::|||   |||      |||||||||   |||:::|||||||||
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu.............SerIleAla 103
                :::::::::|||   :::               :::
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
        :::|||   |||:::      ||||||:::||| 
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349

seq_name: gb_est81:BE970243

seq_documentation_block:
LOCUS       BE970243      772 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679987F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950397 5',
            mRNA sequence.
ACCESSION   BE970243
VERSION     BE970243.1  GI:10583088
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: m  column: 22
            High quality sequence stop: 518.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950397"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
```

(where B = A, C, or G and N = A, C, G, or T). Average insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 208 a 256 c 185 g 123 t
ORIGIN

alignment_scores:
Quality: 135.50 Length: 111
Ratio: 1.908 Gaps: 3
Percent Similarity: 63.964 Percent Identity: 28.829

alignment_block:
US-09-215-435-179 x BE970243 ..

Align seg 1/1 to: BE970243 from: 1 to: 772

```
     12 LeuLeuPhePhePheLeuPheLeuLeuThrArgGlySerLeuSerProTh 28
     26 ATCCTGATCCTCCTGCTTGTCGCCCTCTCTGTGGCCTATGCAGCTCCTGG 75

     28 rLysTyrAsnLeu.........LeuGluLeuLysGluSerCysIleArgA 42
     76 CCCCCGGGGGATCATTATCAACCTGGAGAACGGTGAGCTCTGCATGAATA 125

     42 snGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGlu 58
    126 GTGCCCAGTGTAAGAGCAATTGCTGCCAGCATTCAAGTGCGCTGGGCCTG 175

     59 SerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 75
    176 GCCCGCTGCACATCCATGGCCAGCGAGAACAGCGAGTGCTCTGTCAAGAC 225

     75 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 92
    226 GCTCTATGGGATTTACTACAAGTGTCCCTGTGAGCGTGGCCTGACCTGT. 274

     92 leTyrSerLysAsnGluLysTrpLeu..............SerIleAla 103
    275 ........GAGGGAGACAAGACCATCGTGGGCTCCATCACCAACACCAAC 316

    104 TyrGlyArgCysGlnLysIleGlyArgGlnLys 114
    317 TTTGGCATCTGCCATGACGCTGGACGCTCCAAG 349
```

```
OM of: US-09-215-435-179_COPY_1_23 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:45 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pr6:AL157823         +  119.00  221.87  0.0004    127384  ! AL157823 Human DNA sequence
gb_htg22:AL512371       +  119.00  220.11  0.0005    159482  ! AL512371 Homo sapiens chromo
gb_htg23:AL590400       +  119.00  219.66  0.0005    168884  ! AL590400 Homo sapiens chromo
gb_pr2:AC007056         +   70.00  111.75  548.17    182481  ! AC007056 Homo sapiens clone
gb_htg10:AC026232       +   68.50  108.04  881.75    192561  ! AC026232 Mus musculus chromo
gb_htg14:AC068663       -   68.50  107.26  975.41    212908  ! AC068663 Mus musculus chromo
gb_htg24:AP002376       -   68.00  108.36  846.35    160759  ! AP002376 Homo sapiens chromo
gb_htg18:AC087103       +   68.00  107.85  903.97    171648  ! AC087103 Papio hamadryas cl
gb_htg20:AL158168       +   68.00  107.29  970.60    184233  ! AL158168 Homo sapiens chromo
gb_pr1:AC004644         +   67.00  117.20  272.34     39338  ! AC004644 Homo sapiens chromos
gb_htg25:HSDJ821D9      +   67.00  107.14  989.34    141991  ! AL121960 Homo sapiens chromo
gb_pr8:HS24O18          +   67.00  106.50  1.1e+03    154066  ! AL021808 Human DNA sequence
gb_htg3:AC012220        -   67.00  106.26  1.1e+03    158947  ! AC012220 Homo sapiens chromo
gb_htg2:AC009762        -   67.00  105.29  1.3e+03    179956  ! AC009762 Homo sapiens clone
gb_pr9:HSSDHBS5         +   66.00  147.18  5.83          649  ! U17883 Human succinate dehydrog
gb_htg24:AP003180       +   66.00  104.33  1.4e+03    153814  ! AP003180 Homo sapiens chromo
gb_htg19:AL138896       -   66.00  104.20  1.4e+03    156375  ! AL138896 Homo sapiens chromo
gb_htg24:AP003401       +   66.00  104.07  1.5e+03    158977  ! AP003401 Homo sapiens chromo
gb_htg7:AC021317        +   66.00  103.79  1.5e+03    164847  ! AC021317 Homo sapiens clone
gb_htg7:AC020746        -   66.00  103.70  1.5e+03    166566  ! AC020746 Homo sapiens chromo
gb_htg5:AC015958        +   66.00  103.69  1.5e+03    166960  ! AC015958 Homo sapiens chromo
gb_htg10:AC025335       +   66.00  103.58  1.6e+03    169291  ! AC025335 Homo sapiens chromo
gb_htg19:AL139404       -   66.00  103.50  1.6e+03    170889  ! AL139404 Homo sapiens chromo
gb_htg24:AP003356       -   66.00  103.49  1.6e+03    171154  ! AP003356 Homo sapiens chromo
gb_htg8:AC022245        +   66.00  103.32  1.6e+03    174937  ! AC022245 Homo sapiens chromo
gb_htg14:AC068410       +   66.00  103.20  1.6e+03    177534  ! AC068410 Homo sapiens chromo
gb_htg18:AC090016       +   66.00  103.17  1.6e+03    178360  ! AC090016 Homo sapiens chromo
gb_htg21:AL356142       +   66.00  103.08  1.7e+03    180282  ! AL356142 Homo sapiens chromo
gb_htg22:AL390961       -   66.00  103.02  1.7e+03    181674  ! AL390961 Homo sapiens chromo
gb_htg10:AC026021       +   66.00  102.83  1.7e+03    186248  ! AC026021 Homo sapiens chromo
gb_htg23:AP001095       +   66.00  102.72  1.7e+03    188818  ! AP001095 Homo sapiens chromo
gb_htg14:AC067794       -   66.00  102.62  1.8e+03    191379  ! AC067794 Homo sapiens chromo
gb_htg24:AP002851       +   66.00  102.57  1.8e+03    192465  ! AP002851 Homo sapiens chromo
gb_htg18:AC087138       -   66.00  102.33  1.8e+03    198578  ! AC087138 Mus musculus clone
gb_htg24:AP002756       +   66.00  102.06  1.9e+03    205430  ! AP002756 Homo sapiens chromo
gb_htg13:AC063965       +   66.00  102.01  1.9e+03    206848  ! AC063965 Homo sapiens chromo
gb_htg23:AP001011       -   66.00  101.96  1.9e+03    207949  ! AP001011 Homo sapiens chromo
gb_htg24:AP002757       -   66.00  101.84  2.0e+03    211325  ! AP002757 Homo sapiens chromo
gb_pr9:HSDJ37C10        +   66.00  101.65  2.0e+03    216497  ! AL049569 Human DNA sequence
gb_pr4:AF067844         +   66.00  101.58  2.0e+03    218336  ! AF067844 Homo sapiens chromo
gb_htg6:AC018700        -   66.00  101.49  2.0e+03    220808  ! AC018700 Homo sapiens chromo
gb_in3:DDIRAP1A         -   65.00  136.22  23.77      1987  ! L09750 Dictyostelium discoi
gb_htg4:AC013976        -   65.00  124.07  112.89     9365  ! AC013976 Drosophila melanog
gb_pr8:HS206D15         +   65.00  105.39  1.2e+03    101574  ! AL021068 Homo sapiens DNA
gb_htg19:AL139250       +   65.00  104.98  1.3e+03    106962  ! AL139250 Homo sapiens chr

seq_name: gb_pr6:AL157823

seq_documentation_block:
LOCUS       AL157823   127384 bp    DNA             PRI       10-FEB-2001
DEFINITION  Human DNA sequence from clone RP3-510O8 on chromosome 6, complete
            sequence.
ACCESSION   AL157823
VERSION     AL157823.9  GI:12750800
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127384)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 12, 2001 this sequence version replaced gi:12709869.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-510O8 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-510O8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-510O8 is at 1 in this sequence. The
            true left end of clone RP3-422H11 is at 127285 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .127384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-510O8"
                     /clone_lib="RPCI-3"
     misc_feature    6404
                     /note="Weak data in alu."
     misc_feature    complement(95875. .96037)
                     /note="match: STS: Em:G13580"
     misc_feature    95894. .96129
                     /note="match: STS: Em:G14863"
BASE COUNT    33605 a  30130 c  31408 g  32241 t
ORIGIN
```

```
alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AL157823    ..

 Align seg 1/1  to: AL157823  from: 1  to: 127384

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   82003 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 82052

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
   82053 CTTCCTCCTCACCAGGGGC 82071

seq_name: gb_htg22:AL512371

seq_documentation_block:
LOCUS       AL512371   159482 bp    DNA             HTG       21-MAR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-238F24, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL512371
VERSION     AL512371.4  GI:13274864
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159482)
  AUTHORS   Clark,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 12, 2001 this sequence version replaced gi:12964529.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA238F24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 154970 bases at least Q40
            Consensus quality: 156246 bases at least Q30
            Consensus quality: 156738 bases at least Q20
            Insert size: 159282; sum-of-contigs
            Insert size: 165487; 7.5% error; agarose-fp
            Quality coverage: 6.35x in Q20 bases; sum-of-contigs Quality
            coverage: 6.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3066: contig of 3066 bp in length
            *    3067 3166: gap of      100 bp
            *    3167   155002: contig of 151836 bp in length
            *  155003 155102: gap of      100 bp
            *  155103   159482: contig of 4380 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159482
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-238F24"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .3066
                     /note="assembly_fragment:02384
                     fragment_chain:1"
     misc_feature    3167. .155002
                     /note="assembly_fragment:02553
                     fragment_chain:1"
     misc_feature    155103. .159482
                     /note="assembly_fragment:02886"
BASE COUNT    42530 a  37388 c  38532 g  40831 t    201 others
ORIGIN


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AL512371    ..

 Align seg 1/1  to: AL512371  from: 1  to: 159482

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   95543 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 95592

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
   95593 CTTCCTCCTCACCAGGGGC 95611

seq_name: gb_htg23:AL590400

seq_documentation_block:
LOCUS       AL590400   168884 bp    DNA             HTG       08-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-569I12, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   AL590400
VERSION     AL590400.2  GI:13568356
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168884)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:13559131.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA569I12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 164511 bases at least Q40
            Consensus quality: 165952 bases at least Q30
            Consensus quality: 166791 bases at least Q20
            Insert size: 167584; sum-of-contigs
            Insert size: 185239; 2.7% error; agarose-fp
            Quality coverage: 6.19x in Q20 bases; sum-of-contigs Quality
            coverage: 5.74x in Q20 bases; agarose-fp
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2920: contig of 2920 bp in length
            *    2921 3020: gap of      100 bp
            *    3021    32168: contig of 29148 bp in length
            *   32169 32268: gap of      100 bp
            *   32269    36080: contig of 3812 bp in length
            *   36081 36180: gap of      100 bp
            *   36181    45704: contig of 9524 bp in length
            *   45705 45804: gap of      100 bp
            *   45805    59680: contig of 13876 bp in length
            *   59681 59780: gap of      100 bp
            *   59781    61890: contig of 2110 bp in length
            *   61891 61990: gap of      100 bp
            *   61991    72185: contig of 10195 bp in length
            *   72186 72285: gap of      100 bp
            *   72286    76243: contig of 3958 bp in length
            *   76244 76343: gap of      100 bp
            *   76344    81840: contig of 5497 bp in length
            *   81841 81940: gap of      100 bp
            *   81941    86241: contig of 4301 bp in length
            *   86242 86341: gap of      100 bp
            *   86342   129124: contig of 42783 bp in length
            *  129125 129224: gap of      100 bp
            *  129225   131985: contig of 2761 bp in length
            *  131986 132085: gap of      100 bp
            *  132086   151439: contig of 19354 bp in length
            *  151440 151539: gap of      100 bp
            *  151540   168884: contig of 17345 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-569I12"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .2920
                     /note="assembly_fragment:00514
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3021. .32168
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    32269. .36080
                     /note="assembly_fragment:00777
                     fragment_chain:1"
     misc_feature    36181. .45704
                     /note="assembly_fragment:00971
                     fragment_chain:1"
     misc_feature    45805. .59680
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    59781. .61890
                     /note="assembly_fragment:02150
                     fragment_chain:2"
     misc_feature    61991. .72185
                     /note="assembly_fragment:02579
                     fragment_chain:2"
     misc_feature    72286. .76243
                     /note="assembly_fragment:00659
                     fragment_chain:2"
     misc_feature    76344. .81840
                     /note="assembly_fragment:00425
                     fragment_chain:2"
     misc_feature    81941. .86241
                     /note="assembly_fragment:02764
                     fragment_chain:2"
     misc_feature    86342. .129124
                     /note="assembly_fragment:00846
                     fragment_chain:2"
     misc_feature    129225. .131985
                     /note="assembly_fragment:01963"
     misc_feature    132086. .151439
                     /note="assembly_fragment:02120
                     fragment_chain:3"
     misc_feature    151540. .168884
                     /note="assembly_fragment:01758
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
BASE COUNT    46011 a  38577 c  39446 g  43544 t   1306 others
ORIGIN


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AL590400   ..

 Align seg 1/1  to: AL590400  from: 1  to: 168884

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  109039 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 109088

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
  109089 CTTCCTCCTCACCAGGGGC 109107

seq_name: gb_pr2:AC007056

seq_documentation_block:
LOCUS       AC007056   182481 bp    DNA             PRI       01-JUL-1999
DEFINITION  Homo sapiens clone from human chromosome 14q31 region containing
            gene for neurexin III, partial CDS, complete sequence.
ACCESSION   AC007056
VERSION     AC007056.4  GI:5306227
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182481)
  AUTHORS   Young,J., Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dahl,T.,
            Dickhoff,R., Hall,J., James,R., Loretz,C., Lasky,S., Madan,A.,
            Prescott,S., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Sequencing of human chromosome 14 gene for neurexin III
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182481)
  AUTHORS   Young,J.M., Qin,S., Madan,A., Abbasi,N., Madan,A., Dors,M.,
            James,R., Loretz,C., Lasky,S., Rowen,L. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1999) Multimegabase Sequencing Group, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 182481)
  AUTHORS   Young,J., Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dahl,T.,
            Dickhoff,R., Hall,J., James,R., Loretz,C., Lasky,S., Madan,A.,
            Prescott,S., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Jul 1, 1999 this sequence version replaced gi:4966341.
            This sequence overlaps BAC 332E19, found in GenBank Accession
            Number AF099810.
FEATURES             Location/Qualifiers
     source          1. .182481
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q31"
                     /clone="BAC 31E9"
                     /clone_lib="RPCI-11"
misc_feature         1. .23236
                     /note="Overlap span with BAC 332E19, found in AF099810, at
                     positions 170589-193824."
variation            1702
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            3369
                     /note="31E9: a; 332E19: c"
                     /replace="c"
variation            3867. .3869
                     /note="31E9: ttt; 332E19: t"
                     /replace="t"
variation            4378
                     /note="31E9: g; 332E19: gccag"
                     /replace="gccag"
variation            4858
                     /note="31E9: g; 332E19: a"
                     /replace="a"
variation            5277
                     /note="31E9: g; 332E19: t"
                     /replace="t"
variation            5393
                     /note="31E9: t; 332E19: g"
                     /replace="g"
variation            6561
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            8235
                     /note="31E9: g; 332E19: c"
                     /replace="c"
variation            8591
                     /note="31E9: t; 332E19: a"
                     /replace="a"
variation            10516
                     /note="31E9: a; 332E19: t"
                     /replace="t"
variation            11378
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            13958
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            13994
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            14454
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            14518
                     /note="31E9: g; 332E19: a"
                     /replace="a"
variation            14619
                     /note="31E9: c; 332E19: t"
                     /replace="t"
variation            14679
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            14753
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            15144
                     /note="31E9: t; 332E19: g"
                     /replace="g"
variation            15809
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            16109
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            16639
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            16725
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            17208
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            17311
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            17579
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            17731
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            18114
                     /note="31E9: a; 332E19: c"
                     /replace="c"
variation            18491
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            18728
                     /note="31E9: t; 332E19: c"
                     /replace="c"
variation            19388
                     /note="31E9: g; 332E19: a"
                     /replace="a"
variation            19728
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            20250
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            21522
                     /note="31E9: g; 332E19: a"
                     /replace="a"
variation            21838. .21839
                     /note="31E9: aa; 332E19: a"
                     /replace="a"
variation            21902
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            21978
                     /note="31E9: a; 332E19: c"
                     /replace="c"
variation            22190
                     /note="31E9: c; 332E19: a"
                     /replace="a"
variation            22265
                     /note="31E9: g; 332E19: a"
                     /replace="a"
variation            22355
                     /note="31E9: a; 332E19: g"
                     /replace="g"
variation            22525. .22526
                     /note="31E9: aa; 332E19: a"
                     /replace="a"
variation            22924
                     /note="31E9: t; 332E19: a"
                     /replace="a"
unsure               69442. .69550
                     /note="low quality data"
unsure               69650. .69762
                     /note="low quality data"
CDS                  <75358. .>75426
                     /note="Matches the beginning of KIAA0743, found in
                     AB018286.  The start position of the exon is unclear."
                     /codon_start=3
```

```
                     /product="neurexin III"
                     /protein_id="AAD41968.1"
                     /db_xref="GI:5306228"
                     /translation="CAFSGICTLHFHSCSEINSSLA"
     unsure          91474. .91575
                     /note="low quality data"
     unsure          166633. .166647
                     /note="number of Ts unclear -- could be one fewer"
BASE COUNT    52118 a  35513 c  39070 g  55780 t
ORIGIN


alignment_scores:
            Quality:   70.00            Length:     19
              Ratio:  3.684              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 73.684

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC007056   ..

 Align seg 1/1  to: AC007056  from: 1  to: 182481

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         :::||||||::::::::::||||||||::::|||||||||||||||||||
  128504 TTGCTTCCTTCCTTCTTCCTTCTCCTCCTTCTCCTCTTCTTCTTTCTCTT 128553

      18 eLeuLeu 20
         |||||||
  128554 CCTCCTC 128560

seq_name: gb_htg10:AC026232

seq_documentation_block:
LOCUS       AC026232   192561 bp    DNA             HTG       21-MAR-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-396K2 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 12 unordered pieces.
ACCESSION   AC026232
VERSION     AC026232.1  GI:7272123
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 192561)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Bouffard,G.G.,
            Dietrich,N.L., Eagle,W.O., Gan,W., Gupta,J., Ho,S.-L., Huang,M.C.,
            Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B., Mastrian,S.D.,
            McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R., Stantripop,S.,
            Summers,T.J., Thomas,J.W., Thomas,P.J., Tiongson,E.E.,
            Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
            Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192561)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     ---------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            ---------- Project Information
                Center project name: xj
                Center clone name: 396K02
            ---------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 185640 bases at least Q40
                Consensus quality: 186751 bases at least Q30
                Consensus quality: 187342 bases at least Q20
                Insert size: 181000; agarose-fp
                Insert size: 178000; pulse-field-gel
                Insert size: 192561; sum-of-contigs
                Quality coverage: 8.24x in Q20 bases; agarose-fp
                Quality coverage: 8.38x in Q20 bases; pulse-field-gel
                Quality coverage: 7.74x in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2035: contig of 2035 bp in length
            *                   gap of unknown length
            *     2036     5131: contig of 3096 bp in length
            *                   gap of unknown length
            *     5132     7976: contig of 2845 bp in length
            *                   gap of unknown length
            *     7977    11776: contig of 3800 bp in length
            *                   gap of unknown length
            *    11777    19780: contig of 8004 bp in length
            *                   gap of unknown length
            *    19781    38116: contig of 18336 bp in length
            *                   gap of unknown length
            *    38117    54986: contig of 16870 bp in length
            *                   gap of unknown length
            *    54987    75295: contig of 20309 bp in length
            *                   gap of unknown length
            *    75296   102532: contig of 27237 bp in length
            *                   gap of unknown length
            *   102533   124744: contig of 22212 bp in length
            *                   gap of unknown length
            *   124745   154236: contig of 29492 bp in length
            *                   gap of unknown length
            *   154237   192561: contig of 38325 bp in length.
FEATURES             Location/Qualifiers
     source          1. .192561
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-396K2"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    55008 a  39646 c  39952 g  57733 t    222 others
ORIGIN


alignment_scores:
            Quality:   68.50            Length:     24
              Ratio:  3.605              Gaps:      1
 Percent Similarity:  79.167   Percent Identity: 62.500

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC026232   ..

 Align seg 1/1  to: AC026232  from: 1  to: 192561

       3 LeuProGlnTrp..............LeuLeuLeuLeuPheLeuLeuPh 14
         ||||||:::|||              |||:::|||||||||||||||||
    5171 CTTCCTTCCTGGAATTCATTTCTTCCTCTTTTCCTTCTCTTTCTCCTTTT 5220

      14 ePhePheLeuPheLeuLeuThr 21
         |||||||:::|||:::|||||||
    5221 TTTTTTTTTTTTTTTTTAACT 5242

seq_name: gb_htg14:AC068663

seq_documentation_block:
LOCUS       AC068663   212908 bp    DNA             HTG       01-JUN-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-280D18 strain C57BL6/J,
```

```
            WORKING DRAFT SEQUENCE, 25 unordered pieces.
ACCESSION   AC068663
VERSION     AC068663.2  GI:8134859
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212908)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
            Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
            Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212908)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7712112.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: xi
                 Center clone name: 280D18
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 199975 bases at least Q40
                 Consensus quality: 202777 bases at least Q30
                 Consensus quality: 204027 bases at least Q20
                 Insert size: 222000; agarose-fp
                 Insert size: 216000; pulse-field-gel
                 Insert size: 210508; sum-of-contigs
                 Quality coverage: 4.37x in Q20 bases; agarose-fp
                 Quality coverage: 4.49x in Q20 bases; pulse-field-gel
                 Quality coverage: 4.61x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2371: contig of 2371 bp in length
            *     2372     2471: gap of unknown length
            *     2472     4622: contig of 2151 bp in length
            *     4623     4722: gap of unknown length
            *     4723     7916: contig of 3194 bp in length
            *     7917     8016: gap of unknown length
            *     8017    10417: contig of 2401 bp in length
            *    10418    10517: gap of unknown length
            *    10518    14260: contig of 3743 bp in length
            *    14261    14360: gap of unknown length
            *    14361    18020: contig of 3660 bp in length
            *    18021    18120: gap of unknown length
            *    18121    22445: contig of 4325 bp in length
            *    22446    22545: gap of unknown length
            *    22546    26537: contig of 3992 bp in length
            *    26538    26637: gap of unknown length
            *    26638    30194: contig of 3557 bp in length
            *    30195    30294: gap of unknown length
            *    30295    34805: contig of 4511 bp in length
            *    34806    34905: gap of unknown length
            *    34906    39674: contig of 4769 bp in length
            *    39675    39774: gap of unknown length
            *    39775    44570: contig of 4796 bp in length
            *    44571    44670: gap of unknown length
            *    44671    49969: contig of 5299 bp in length
            *    49970    50069: gap of unknown length
            *    50070    55625: contig of 5556 bp in length
            *    55626    55725: gap of unknown length
            *    55726    61502: contig of 5777 bp in length
            *    61503    61602: gap of unknown length
            *    61603    70931: contig of 9329 bp in length
            *    70932    71031: gap of unknown length
            *    71032    79217: contig of 8186 bp in length
            *    79218    79317: gap of unknown length
            *    79318    88380: contig of 9063 bp in length
            *    88381    88480: gap of unknown length
            *    88481   102624: contig of 14144 bp in length
            *   102625   102724: gap of unknown length
            *   102725   112658: contig of 9934 bp in length
            *   112659   112758: gap of unknown length
            *   112759   130169: contig of 17411 bp in length
            *   130170   130269: gap of unknown length
            *   130270   149318: contig of 19049 bp in length
            *   149319   149418: gap of unknown length
            *   149419   170076: contig of 20658 bp in length
            *   170077   170176: gap of unknown length
            *   170177   189732: contig of 19556 bp in length
            *   189733   189832: gap of unknown length
            *   189833   212908: contig of 23076 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212908
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-280D18"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2371
                     /note="assembly_fragment"
     misc_feature    2472. .4622
                     /note="assembly_fragment"
     misc_feature    4723. .7916
                     /note="assembly_fragment"
     misc_feature    8017. .10417
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    10518. .14260
                     /note="assembly_fragment"
     misc_feature    14361. .18020
                     /note="assembly_fragment"
     misc_feature    18121. .22445
                     /note="assembly_fragment"
     misc_feature    22546. .26537
                     /note="assembly_fragment"
     misc_feature    26638. .30194
                     /note="assembly_fragment"
     misc_feature    30295. .34805
                     /note="assembly_fragment"
     misc_feature    34906. .39674
                     /note="assembly_fragment"
     misc_feature    39775. .44570
                     /note="assembly_fragment"
     misc_feature    44671. .49969
                     /note="assembly_fragment"
     misc_feature    50070. .55625
                     /note="assembly_fragment"
     misc_feature    55726. .61502
                     /note="assembly_fragment"
     misc_feature    61603. .70931
                     /note="assembly_fragment"
     misc_feature    71032. .79217
```

```
                     /note="assembly_fragment"
     misc_feature    79318. .88380
                     /note="assembly_fragment"
     misc_feature    88481. .102624
                     /note="assembly_fragment"
     misc_feature    102725. .112658
                     /note="assembly_fragment"
     misc_feature    112759. .130169
                     /note="assembly_fragment"
     misc_feature    130270. .149318
                     /note="assembly_fragment"
     misc_feature    149419. .170076
                     /note="assembly_fragment"
     misc_feature    170177. .189732
                     /note="assembly_fragment"
     misc_feature    189833. .212908
                     /note="assembly_fragment"
BASE COUNT    61004 a  43094 c  43404 g  62838 t   2568 others
ORIGIN


alignment_scores:
            Quality:   68.50              Length:     24
              Ratio:  3.605               Gaps:        1
 Percent Similarity:  79.167   Percent Identity: 62.500

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC068663/rev   ..

 Align seg 1/1  to reverse of: AC068663  from: 1  to: 212908

       3 LeuProGlnTrp.............LeuLeuLeuLeuPheLeuLeuPh 14
         ||||||:::|||              |||:::||||||||||||||||||
   61387 CTTCCTTCCTGGCATTCATTTCTTCCTCTTTTCCTTCTCTTTCTCCTTTT 61338

      14 ePhePheLeuPheLeuLeuThr 21
         |||||||:::|||:::||||||
   61337 TTTTTTTTTTTTTTTTTTAACT 61316

seq_name: gb_htg24:AP002376

seq_documentation_block:
LOCUS       AP002376   160759 bp    DNA             HTG       03-NOV-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-349I16 map 11q, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AP002376
VERSION     AP002376.2  GI:11094148
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-349I16.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160759)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 160,759 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 160759)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Nov 3, 2000 this sequence version replaced gi:8131640.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-349I16
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 157952 bases at least Q40
                Consensus quality: 159001 bases at least Q30
                Consensus quality: 159356 bases at least Q20
                Insert size: 159859; sum-of-contigs
                Quality coverage: 9.52x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            10 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     48185 contig of  48185 bp in length
              48286     89029 contig of  40744 bp in length
              89130    107881 contig of  18752 bp in length
             107982    125922 contig of  17941 bp in length
             126023    137453 contig of  11431 bp in length
             137554    145908 contig of   8355 bp in length
             146009    151846 contig of   5838 bp in length
             151947    156237 contig of   4291 bp in length
             156338    159336 contig of   2999 bp in length
             159437    160759 contig of   1323 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    48185: contig of 48185 bp in length
            *   48186 48285: gap of      100 bp
            *   48286    89029: contig of 40744 bp in length
            *   89030 89129: gap of      100 bp
            *   89130   107881: contig of 18752 bp in length
            *  107882 107981: gap of      100 bp
            *  107982   125922: contig of 17941 bp in length
            *  125923 126022: gap of      100 bp
            *  126023   137453: contig of 11431 bp in length
            *  137454 137553: gap of      100 bp
            *  137554   145908: contig of 8355 bp in length
            *  145909 146008: gap of      100 bp
            *  146009   151846: contig of 5838 bp in length
            *  151847 151946: gap of      100 bp
            *  151947   156237: contig of 4291 bp in length
            *  156238 156337: gap of      100 bp
            *  156338   159336: contig of 2999 bp in length
            *  159337 159436: gap of      100 bp
            *  159437   160759: contig of 1323 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160759
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-349I16"
     misc_feature    1. .48185
                     /note="assembly_fragment"
     misc_feature    48286. .89029
                     /note="assembly_fragment"
     misc_feature    89130. .107881
                     /note="assembly_fragment"
     misc_feature    107982. .125922
                     /note="assembly_fragment"
     misc_feature    126023. .137453
                     /note="assembly_fragment"
```

```
     misc_feature    137554. .145908
                     /note="assembly_fragment"
     misc_feature    146009. .151846
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    151947. .156237
                     /note="assembly_fragment"
     misc_feature    156338. .159336
                     /note="assembly_fragment"
     misc_feature    159437. .160759
                     /note="assembly_fragment clone_end:T7 vector_side:right"
BASE COUNT    46360 a  32519 c  33059 g  47918 t    903 others
ORIGIN


alignment_scores:
            Quality:   68.00              Length:     18
              Ratio:  4.533                Gaps:      0
 Percent Similarity:  83.333   Percent Identity: 72.222

alignment_block:
 US-09-215-435-179_COPY_1_23 x AP002376/rev   ..

 Align seg 1/1  to reverse of: AP002376  from: 1  to: 160759

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         :::|||   ||||||||||||||||||   :::||||||||||   |||||
   90328 ATCCCTGTGTGGTTGTTACTTCTTTTTTCTTTTTTCTTTTTTTCTTTTTT 90279

      19 uLeu 20
         ||||
   90278 ATTA 90275

 seq_name: gb_htg18:AC087103

 seq_documentation_block:
 LOCUS       AC087103   171648 bp    DNA             HTG       07-DEC-2000
 DEFINITION  Papio hamadryas clone RP41-353B7, WORKING DRAFT SEQUENCE, 6
             unordered pieces.
 ACCESSION   AC087103
 VERSION     AC087103.1  GI:11596994
 KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE      baboon.
   ORGANISM  Papio hamadryas
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
             Cercopithecinae; Papio.
 REFERENCE   1  (bases 1 to 171648)
   AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
             Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
             Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
             Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
             Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
             Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
             Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
             Wetherby,K.D., Zhang,L.-H. and Green,E.D.
   TITLE     NISC Comparative Sequencing Initiative
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 171648)
   AUTHORS   Green,E.D.
   TITLE     Direct Submission
   JOURNAL   Submitted (07-DEC-2000) NIH Intramural Sequencing Center, 8717
             Grovemont Circle, Gaithersburg, MD 20877, USA
 COMMENT     -------------- Genome Center
                  Center: NIH Intramural Sequencing Center
                  Center code: NISC
                  Web site: http://www.nisc.nih.gov
                  Contact: nisc_mouse@nhgri.nih.gov
             -------------- Project Information
                  Center project name: asb
                  Center clone name: 353B07
             -------------- Summary Statistics
                  Sequencing vector: plasmid; n/a; 100% of reads
                  Chemistry: Dye-terminator Big Dye; 100% of reads
                  Assembly program: Phrap; version 0.990319
                  Consensus quality: 168876 bases at least Q40
                  Consensus quality: 169963 bases at least Q30
                  Consensus quality: 170362 bases at least Q20
                  Insert size: 145000; agarose-fp
                  Insert size: 171148; sum-of-contigs
                  Quality coverage: 9.21x in Q20 bases; agarose-fp
                  Quality coverage: 7.81x in Q20 bases; sum-of-contigs
             --------------.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 6 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *      1    11948: contig of 11948 bp in length
             *  11949    12048: gap of unknown length
             *  12049    25411: contig of 13363 bp in length
             *  25412    25511: gap of unknown length
             *  25512    48565: contig of 23054 bp in length
             *  48566    48665: gap of unknown length
             *  48666    78593: contig of 29928 bp in length
             *  78594    78693: gap of unknown length
             *  78694   125203: contig of 46510 bp in length
             * 125204   125303: gap of unknown length
             * 125304   171648: contig of 46345 bp in length.
 FEATURES             Location/Qualifiers
      source          1. .171648
                      /organism="Papio hamadryas"
                      /db_xref="taxon:9557"
                      /clone="RP41-353B7"
                      /clone_lib="RP41"
      misc_feature    1. .11948
                      /note="assembly_fragment"
      misc_feature    12049. .25411
                      /note="assembly_fragment
                      clone_end:T7
                      vector_side:right"
      misc_feature    25512. .48565
                      /note="assembly_fragment"
      misc_feature    48666. .78593
                      /note="assembly_fragment"
      misc_feature    78694. .125203
                      /note="assembly_fragment"
      misc_feature    125304. .171648
                      /note="assembly_fragment
                      clone_end:SP6
                      vector_side:right"
 BASE COUNT    53294 a  34223 c  31880 g  51736 t    515 others
 ORIGIN


 alignment_scores:
             Quality:   68.00              Length:     19
               Ratio:  3.778                Gaps:      0
  Percent Similarity:  94.737   Percent Identity: 57.895

 alignment_block:
  US-09-215-435-179_COPY_1_23 x AC087103   ..

  Align seg 1/1  to: AC087103  from: 1  to: 171648

        1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
          ::::::||||||   |||:::|||||||:::|||::::::||||||||::
    29319 TTACTGCTGCCTGTGTGGTTTCTGTTAATTTTTTTTTTTTTTTTTTTTT 29368

       17 uPheLeu 19
          :|||||
    29369 TTTTTTG 29375

 seq_name: gb_htg20:AL158168
```

```
seq_documentation_block:
LOCUS       AL158168   184233 bp    DNA             HTG       07-APR-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-7C6, *** SEQUENCING IN
            PROGRESS ***, 5 unordered pieces.
ACCESSION   AL158168
VERSION     AL158168.18  GI:13561142
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184233)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 8, 2001 this sequence version replaced gi:13274297.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA7C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 182545 bases at least Q40
            Consensus quality: 183275 bases at least Q30
            Consensus quality: 183594 bases at least Q20
            Insert size: 183833; sum-of-contigs
            Insert size: 141612; 18.6% error; agarose-fp
            Quality coverage: 6.90x in Q20 bases; sum-of-contigs Quality
            coverage: 9.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    53357: contig of 53357 bp in length
            *    53358 53457: gap of      100 bp
            *    53458    70030: contig of 16573 bp in length
            *    70031 70130: gap of      100 bp
            *    70131   101423: contig of 31293 bp in length
            *   101424 101523: gap of      100 bp
            *   101524   104203: contig of 2680 bp in length
            *   104204 104303: gap of      100 bp
            *   104304   184233: contig of 79930 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-7C6"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .53357
                     /note="assembly_fragment:03094
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    53458. .70030
                     /note="assembly_fragment:02498
                     fragment_chain:1"
     misc_feature    70131. .101423
                     /note="assembly_fragment:02014
                     fragment_chain:1"
     misc_feature    101524. .104203
                     /note="assembly_fragment:01421"
     misc_feature    104304. .184233
                     /note="assembly_fragment:02355"
BASE COUNT    52765 a  39531 c  38779 g  52756 t    402 others
ORIGIN


alignment_scores:
            Quality:   68.00              Length:     20
              Ratio:  3.778                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AL158168   ..
   .     .
 Align seg 1/1  to: AL158168  from: 1  to: 184233

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         ||||||:::::||||||   :::|||||||||||||||||||||||||||
  101200 CTTCCTTCCTTCCTTTTGTCTTTCTTTCTTCTTTTTTTCTTTCTATTTCT 101249

      19 uLeuThrArg 22
         |:::   |||
  101250 TTTTTTGAGG 101259

seq_name: gb_pri:AC004644

seq_documentation_block:
LOCUS       AC004644    39338 bp    DNA             PRI       01-MAY-1998
DEFINITION  Homo sapiens chromosome 16, cosmid clone 367E12 (LANL), complete
            sequence.
ACCESSION   AC004644
VERSION     AC004644.1  GI:3097840
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39338)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Sequencing of Human Chromosome 16p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39338)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39338)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1. .39338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="367E12"
                     /chromosome="16"
                     /map="16p13.3"
     repeat_region   891. .1355
                     /rpt_family="Alu"
     repeat_region   2173. .2522
                     /rpt_family="Alu"
```

```
repeat_region   complement(4172. .4490)
                /rpt_family="Alu"
repeat_region   4865. .5171
                /rpt_family="Alu"
repeat_region   5115. .5147
                /note="(A)33"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   5175. .5317
                /rpt_family="Alu"
repeat_region   complement(5713. .6019)
                /rpt_family="Alu"
repeat_region   6123. .6142
                /note="(T)20"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   complement(6123. .6418)
                /rpt_family="Alu"
repeat_region   6567. .6850
                /rpt_family="Alu"
repeat_region   complement(7198. .7496)
                /rpt_family="Alu"
repeat_region   complement(7587. .7804)
                /rpt_family="Alu"
misc_feature    complement(8691. .8782)
                /note="GRAIL 2 excellent exon, frame 1"
repeat_region   8863. .8967
                /rpt_family="Alu"
repeat_region   9063. .9153
                /rpt_family="MIR"
repeat_region   9403. .9757
                /rpt_family="Alu"
repeat_region   9733. .9757
                /note="(A)25"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   complement(9825. .10147)
                /rpt_family="Alu"
repeat_region   complement(10413. .10730)
                /rpt_family="Alu"
repeat_region   11034. .11292
                /rpt_family="Alu"
repeat_region   complement(11360. .11685)
                /rpt_family="Alu"
repeat_region   11360. .11381
                /note="(T)22"
                /rpt_type=tandem
                /rpt_unit=T
misc_feature    11701. .11949
                /note="100% identity EST ab01b06.s1"
                /db_xref="dbEST:AA491442"
repeat_region   complement(12256. .12850)
                /rpt_family="Alu"
repeat_region   complement(13025. .13392)
                /rpt_family="Alu"
repeat_region   13025. .13051
                /note="(T)27"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   13509. .13648
                /rpt_family="L1"
repeat_region   13656. .13682
                /note="(T)27"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   complement(13762. .13864)
                /rpt_family="MIR"
misc_feature    14267. .16252
                /note="GRAIL 2 excellent exon, frame 1"
prim_transcript complement(15185. .15716)
                /standard_name="fmfc16"
                /note="96% identity dbEST:AA631949"
repeat_region   complement(16226. .16423)
                /rpt_family="MLT1"
repeat_region   complement(16476. .16818)
                /rpt_family="Alu"
misc_feature    17364. .17450
                /note="GRAIL 2 excellent exon, frame 0"
repeat_region   complement(17809. .18097)
                /rpt_family="Alu"
repeat_region   18194. .18517
                /rpt_family="Alu"
repeat_region   complement(18699. .19020)
                /rpt_family="Alu"
repeat_region   19197. .19533
                /rpt_family="Alu"
repeat_region   19724. .19786
                /rpt_family="Alu"
repeat_region   20071. .20377
                /rpt_family="Alu"
repeat_region   20544. .20569
                /note="(TG)13"
                /rpt_type=tandem
                /rpt_unit=TG
repeat_region   complement(20796. .21116)
                /rpt_family="Alu"
misc_feature    22158. .22241
                /note="GRAIL 2 excellent exon, frame 0"
repeat_region   22312. .22657
                /rpt_family="Alu"
repeat_region   complement(22798. .23106)
                /rpt_family="Alu"
repeat_region   22798. .22833
                /note="(T)36"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   23181. .23228
                /note="(ATTT)12"
                /rpt_type=tandem
                /rpt_unit=ATTT
repeat_region   23236. .23571
                /rpt_family="Alu"
repeat_region   complement(23575. .23897)
                /rpt_family="L1MB7"
repeat_region   complement(23893. .24208)
                /rpt_family="Alu"
repeat_region   complement(24207. .24547)
                /rpt_family="L1"
misc_feature    24827. .24910
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   25089. .25441
                /rpt_family="Alu"
repeat_region   complement(25787. .25965)
                /rpt_family="MER67C"
repeat_region   complement(26193. .26492)
                /rpt_family="Alu"
misc_feature    26873. .26956
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   27128. .27171
                /note="(ATGG)11"
                /rpt_type=tandem
                /rpt_unit=ATGG
repeat_region   27534. .27561
                /note="(ATGG)7"
                /rpt_type=tandem
                /rpt_unit=ATGG
repeat_region   27882. .27950
                /note="75% identity"
                /rpt_type=tandem
                /rpt_unit=TGAA
misc_feature    28667. .28750
                /note="GRAIL 2 excellent exon, frame 2"
misc_feature    29384. .29467
                /note="GRAIL 2 excellent exon, frame 2"
misc_feature    29701. .29799
                /note="96% identity EST nj24c09.s1"
```

```
                     /db_xref="dbEST:AA564923"
     misc_feature    complement(29935. .30013)
                     /note="84% identity mouse EST mq96b12.r1"
                     /db_xref="dbEST:AA138320"
     misc_feature    29939. .30109
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    30411. .30498
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    30946. .31029
                     /note="GRAIL 2 excellent exon, frame 1"

alignment_scores:
            Quality:   67.00              Length:     17
              Ratio:  4.188                Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 70.588

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC004644   ..

 Align seg 1/1  to: AC004644  from: 1  to: 39338

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         |||:::|||||||||||||   |||::::::|||||||||:::|||||||||
   13636 CCAGAGTGGCTGCTGCTGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTGCT 13685

      20 u 20
         |
   13686 T 13686

seq_name: gb_htg25:HSDJ821D9

seq_documentation_block:
LOCUS       HSDJ821D9  141991 bp    DNA             HTG       24-JAN-2001
DEFINITION  Homo sapiens chromosome 6 clone RP5-821D9 map p21.33-22.3, ***
            SEQUENCING IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AL121960
VERSION     AL121960.4  GI:9885194
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141991)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 22, 2000 this sequence version replaced gi:9864405.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ821D9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 38% of reads
            Sequencing vector: plasmid; L08752; 61% of reads
            Chemistry: Dye-terminator ABI; 10% of reads
            Chemistry: Dye-terminator Big Dye; 73% of reads
            Chemistry: Dye-primer-amersham; 16% of reads
            Consensus quality: 139067 bases at least Q40
            Consensus quality: 139932 bases at least Q30
            Consensus quality: 140472 bases at least Q20
            Insert size: 141291; sum-of-contigs
            Insert size: 129748; 10.1% error; agarose-fp
            Quality coverage: 6.03x in Q20 bases; sum-of-contigs Quality
            coverage: 6.59x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8163: contig of 8163 bp in length
            *     8164 8263: gap of      100 bp
            *     8264    41141: contig of 32878 bp in length
            *    41142 41241: gap of      100 bp
            *    41242    53376: contig of 12135 bp in length
            *    53377 53476: gap of      100 bp
            *    53477    69941: contig of 16465 bp in length
            *    69942 70041: gap of      100 bp
            *    70042    89315: contig of 19274 bp in length
            *    89316 89415: gap of      100 bp
            *    89416    93350: contig of 3935 bp in length
            *    93351 93450: gap of      100 bp
            *    93451   120437: contig of 26987 bp in length
            *   120438 120537: gap of      100 bp
            *   120538   141991: contig of 21454 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p21.33-22.3"
                     /clone="RP5-821D9"
                     /clone_lib="RPCI-5"
     misc_feature    1. .8163
                     /note="assembly_fragment:00030
                     fragment_chain:1"
     misc_feature    8264. .41141
                     /note="assembly_fragment:02333
                     fragment_chain:1"
     misc_feature    41242. .53376
                     /note="assembly_fragment:01379
                     fragment_chain:2"
     misc_feature    53477. .69941
                     /note="assembly_fragment:00831
                     fragment_chain:2"
     misc_feature    70042. .89315
                     /note="assembly_fragment:00126"
     misc_feature    89416. .93350
                     /note="assembly_fragment:00419"
     misc_feature    93451. .120437
                     /note="assembly_fragment:00857"
     misc_feature    120538. .141991
                     /note="assembly_fragment:01760.0"
BASE COUNT    40025 a  29960 c  30808 g  40498 t    700 others
ORIGIN


alignment_scores:
            Quality:   67.00              Length:     22
              Ratio:  3.526                Gaps:      0
 Percent Similarity:  86.364   Percent Identity: 54.545

alignment_block:
 US-09-215-435-179_COPY_1_23 x HSDJ821D9   ..

 Align seg 1/1  to: HSDJ821D9  from: 1  to: 141991

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||:::|||   :::|||:::|||:::|||||||||:::|||   |||::
   14064 ATGGTCCTGGGCAGATGGGTTTTATTTTTGTTTTTGTTTTTTGTATTTAT 14113

      17 uPheLeuLeuThrArg 22
         ::::||||||   |||
   14114 TTATTTATTGAGACGG 14129

seq_name: gb_pr8:HS24018
```

```
seq_documentation_block:
LOCUS       HS24O18    154066 bp    DNA             PRI       15-JUN-2000
DEFINITION  Human DNA sequence from clone XXbac-24o18 on chromosome
            6p21.31-22.2 Contains zinc finger protein pseudogene, VNO-type
            olfactory receptor pseudogene, the gene for PRSS16 (protease,
            serine, 16 (thymus)), the gene for nuclear envelope pore membrane
            protein, ESTs, STSs, and GSSs, complete sequence.
ACCESSION   AL021808
VERSION     AL021808.1  GI:3395513
KEYWORDS    HTG; nuclear envelope pore membrane protein; olfactory receptor;
            protease; PRSS16; zinc finger.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154066)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 5, 1998 this sequence version replaced gi:3355607.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            XXbac-24o18 is from the Research Genetics total human BAC library
            that was screened by D. Ruddy and A. Gnirke from Mercator Genetics
            Inc. as published in Lauer et al (1997) Gen Res. 7: 457-470. This
            sequence is the entire insert of clone XXbac-24o18.
FEATURES             Location/Qualifiers
     source          1. .154066
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p21.31-22.2"
                     /clone_lib="SCb"
                     /clone="XXbac-24o18"
     repeat_region   1. .1043
                     /note="MER52A repeat: matches 217. .1387 of consensus"
     repeat_region   7429. .7472
                     /note="22 copies 2 mer tt 77% conserved"
     misc_feature    complement(7815. .8187)
                     /note="match: GSS: Em:AQ110321"
     misc_feature    8219. .8655
                     /note="match: GSS: Em:AQ338654"
     gene            8618. .11014
                     /gene="b24o18.1"
     CDS             <8618. .>11014
                     /gene="b24o18.1"
                     /note="b24o18.1 (zinc finger)
                     match: cDNAs: Em:U66561 Em:L11672 Em:AF033199 Em:AF024693
                     Em:X59244 Em:AF020591 Em:X74855 Em:M27878 Em:D31763
                     Em:AF011573 Em:U75454 Em:M55422 Em:AF116030 Em:D89928
                     Em:M29580 Em:X17617 Em:U35376 Em:D70831 Em:X78925
                     Em:AB018341 Em:AF086368 Em:Y00850 Em:AB014528 Em:AB007872
                     Em:X65233 Em:M21977 Em:U78722 Em:L77247 Em:U09413
                     Em:X78933 Em:U38904 Em:L32163 Em:X78926 Em:U71599
                     Em:X84801 Em:U07861
                     match: ESTs: Em:T40438 Em:AI650923 Em:T29508 Em:T39154
                     Em:Aa083845 Em:AA716653 Em:AA635575 Em:AA635567
                     Em:AI453500 Em:T90597 Em:D31299
                     match: proteins: Sw:P08045 Tr:O60792 Sw:Q99676 Sw:O75820
                     Tr:O75802 Sw:Q16587 Sw:P51523 Tr:Q14587 Tr:Q62512
                     Sw:P51522 Tr:O43345 Sw:Q03923 Tr:Q14585 Tr:O60765
                     Sw:Q07230 Sw:Q61751 Sw:P52742 Sw:P18753 Tr:Q62886
                     Sw:P13682 Sw:P51786 Sw:P52738 Sw:P18739 Sw:Q62981
                     Tr:Q61116 Tr:O14845 Tr:O75123 Tr:P70591 Sw:P51814
                     Sw:Q16600 Tr:Q15776 Sw:P10076 Tr:Q14586 Tr:Q13398
                     Sw:Q02975 Wp:CE16599 Sw:O43830 Sw:Q62513 Tr:O43339
                     Sw:P21506 Tr:O43167 Sw:P15822 Sw:P51508 Sw:P18751
                     Sw:Q14929"
                     /pseudo
                     /codon_start=1
                     /evidence=not_experimental
     misc_feature    complement(8897. .9335)
                     /note="match: GSS: Em:AQ423094"
     misc_feature    complement(8977. .9310)
                     /note="match: GSS: Em:AQ461858"
     misc_feature    complement(9925. .9966)
                     /note="match: GSS: Em:AQ793807"
     misc_feature    10443. .10483
                     /gene="b24o18.1"
                     /note="match: GSS: Em:AQ376888"
     misc_feature    complement(10536. .10666)
                     /note="match: GSS: Em:U77516"
     misc_feature    complement(10689. .11016)
                     /note="match: STS: Em:G11374"
     misc_feature    complement(10854. .11016)
                     /note="match: STS: Em:G14684"
     misc_feature    complement(11017. .11139)
                     /note="match: GSS: Em:U94452"
     repeat_region   12877. .13020
                     /note="L1MA4A repeat: matches 6119. .6280 of consensus"
     repeat_region   13021. .13761
                     /note="Tigger3(Golem) repeat: matches 2277. .3027 of
                     consensus"
     repeat_region   13753. .13859
                     /note="Tigger3(Golem) repeat: matches 2. .108 of consensus"
     repeat_region   13860. .15869
                     /note="L1MA4A repeat: matches 4267. .6122 of consensus"
     repeat_region   15870. .16491
                     /note="LTR14B repeat: matches 5. .608 of consensus"
     repeat_region   16492. .19102
                     /note="L1MA4A repeat: matches 1857. :4267 of consensus"
     repeat_region   19402. .21319
                     /note="L1MA4A repeat: matches -675. .1857 of consensus"
     misc_feature    22197. .22580
                     /note="match: GSS: Em:AQ799446"
     repeat_region   22697. .22798
                     /note="51 copies 2 mer at 78% conserved"
     misc_feature    complement(23267. .23728)
                     /note="match: GSS: Em:AQ554556"
     misc_feature    complement(23715. .24097)
                     /note="match: GSS: Em:AQ266585"
     repeat_region   25656. .26249
                     /note="L1ME3 repeat: matches 4489. .5132 of consensus"
     repeat_region   26250. .26531
                     /note="AluSc repeat: matches 1. .282 of consensus"
     repeat_region   26532. .26785
                     /note="L1ME3 repeat: matches 5132. .5374 of consensus"
     repeat_region   27386. .27650
                     /note="MLT1G repeat: matches 10. .269 of consensus"
     repeat_region   27671. .27793
                     /note="L1ME3 repeat: matches 5403. .5511 of consensus"
     repeat_region   28078. .28455
                     /note="L1ME3 repeat: matches 5511. .5950 of consensus"
     misc_feature    29086. .29507
                     /note="match: GSS: Em:AQ033258"
```

```
     repeat_region   31337. .31493
                     /note="MER69 repeat: matches 167. .351 of consensus"
     repeat_region   32074. .32140
                     /note="tRNA-Met-i repeat: matches 1. .66 of consensus"
     repeat_region   32262. .32358
                     /note="Charlie2 repeat: matches 254. .349 of consensus"
     repeat_region   32364. .32443
                     /note="Charlie2 repeat: matches 825. .913 of consensus"
     repeat_region   32447. .32908
                     /note="Charlie2 repeat: matches 3193. .3728 of consensus"
     repeat_region   33417. .33482
                     /note="33 copies 2 mer tt 69% conserved"
     repeat_region   33777. .33850
                     /note="tRNA-Lys-AAA repeat: matches 1. .74 of consensus"
     misc_feature    34254. .34507
                     /note="match: STS: Em:G31231"
     repeat_region   34825. .34912
                     /note="L1MA5 repeat: matches 6204. .6293 of consensus"
     repeat_region   34919. .35369
                     /note="Charlie1 repeat: matches 1585. .2053 of consensus"
     repeat_region   35783. .35854
                     /note="tRNA-Met-i repeat: matches 1. .72 of consensus"
     repeat_region   36461. .36494
                     /note="17 copies 2 mer aa 91% conserved"
     misc_feature    36567. .36773
                     /note="match: GSS: Em:AQ382899"
     repeat_region   36700. .36750
                     /note="tRNA-Lys-AAA repeat: matches 8. .58 of consensus"
     repeat_region   37278. .37700
                     /note="MLT1F repeat: matches 3. .541 of consensus"
     repeat_region   38187. .38699
                     /note="L1ME3A repeat: matches 5629. .6154 of consensus"
     repeat_region   39336. .39440
                     /note="L1ME3A repeat: matches 6052. .6156 of consensus"
     repeat_region   41138. .41871
                     /note="MER49 repeat: matches 213. .923 of consensus"
     repeat_region   41872. .41895
                     /note="12 copies 2 mer tc 100% conserved"
     repeat_region   42057. .42179
                     /note="MER49 repeat: matches 100. .222 of consensus"
     gene            complement(43071. .43945)
                     /gene="b24o18.2"
     CDS             complement(43071. .43945)
                     /gene="b24o18.2"
                     /note="b24o18.2 (hs6V1-1)
                     match: cDNAs: Em:U36786
                     match: proteins: Tr:Q62854 Tr:Q62855 Tr:Q62852 Tr:Q62850
                     Tr:Q62853 Tr:Q62856 Tr:Q62851"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
     misc_feature    complement(43630. .44080)
                     /note="match: GSS: Em:AQ704141"
     repeat_region   44882. .44929
                     /note="MLT2FB repeat: matches 367. .414 of consensus"

alignment_scores:
            Quality:   67.00            Length:     22
              Ratio:  3.526              Gaps:      0
 Percent Similarity:  86.364   Percent Identity: 54.545

alignment_block:
 US-09-215-435-179_COPY_1_23 x HS24O18   ..

 Align seg 1/1  to: HS24O18  from: 1  to: 154066

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||:::|||   :::|||:::|||:::||||||||||:::|||   |||::
  108525 ATGGTCCTGGGCAGATGGGTTTTATTTTTGTTTTTGTTTTTTGTATTTAT 108574

      17 uPheLeuLeuThrArg 22
         ::::||||||   |||
  108575 TTATTTATTGAGACGG 108590
```

```
seq_name: gb_htg3:AC012220

seq_documentation_block:
LOCUS       AC012220   158947 bp    DNA             HTG       25-MAY-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-270G15 map 3, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AC012220
VERSION     AC012220.5  GI:8072476
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158947)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-270G15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158947)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7712147.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1822
                Center clone name: 270_G_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 143649 bases at least Q40
                Consensus quality: 150959 bases at least Q30
                Consensus quality: 153947 bases at least Q20
                Insert size: 168000; agarose-fp
                Insert size: 156547; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1062: contig of 1062 bp in length
            *     1063 1162: gap of      100 bp
            *     1163     2772: contig of 1610 bp in length
            *     2773 2872: gap of      100 bp
            *     2873     4182: contig of 1310 bp in length
```

```
            *     4183 4282: gap of      100 bp
            *     4283    6854: contig of 2572 bp in length
            *     6855 6954: gap of      100 bp
            *     6955    9109: contig of 2155 bp in length
            *     9110 9209: gap of      100 bp
            *     9210   10726: contig of 1517 bp in length
            *    10727 10826: gap of      100 bp
            *    10827   13819: contig of 2993 bp in length
            *    13820 13919: gap of      100 bp
            *    13920   17308: contig of 3389 bp in length
            *    17309 17408: gap of      100 bp
            *    17409   21780: contig of 4372 bp in length
            *    21781 21880: gap of      100 bp
            *    21881   26334: contig of 4454 bp in length
            *    26335 26434: gap of      100 bp
            *    26435   31916: contig of 5482 bp in length
            *    31917 32016: gap of      100 bp
            *    32017   38192: contig of 6176 bp in length
            *    38193 38292: gap of      100 bp
            *    38293   43903: contig of 5611 bp in length
            *    43904 44003: gap of      100 bp
            *    44004   48720: contig of 4717 bp in length
            *    48721 48820: gap of      100 bp
            *    48821   54929: contig of 6109 bp in length
            *    54930 55029: gap of      100 bp
            *    55030   61995: contig of 6966 bp in length
            *    61996 62095: gap of      100 bp
            *    62096   68368: contig of 6273 bp in length
            *    68369 68468: gap of      100 bp
            *    68469   73768: contig of 5300 bp in length
            *    73769 73868: gap of      100 bp
            *    73869   81655: contig of 7787 bp in length
            *    81656 81755: gap of      100 bp
            *    81756   93898: contig of 12143 bp in length
            *    93899 93998: gap of      100 bp
            *    93999  103949: contig of 9951 bp in length
            *   103950 104049: gap of      100 bp
            *   104050  115686: contig of 11637 bp in length
            *   115687 115786: gap of      100 bp
            *   115787  128950: contig of 13164 bp in length
            *   128951 129050: gap of      100 bp
            *   129051  145870: contig of 16820 bp in length
            *   145871 145970: gap of      100 bp
            *   145971  158947: contig of 12977 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-270G15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1062
                     /note="assembly_fragment"
     misc_feature    1163. .2772
                     /note="assembly_fragment"
     misc_feature    2873. .4182
                     /note="assembly_fragment"
     misc_feature    4283. .6854
                     /note="assembly_fragment"
     misc_feature    6955. .9109
                     /note="assembly_fragment"
     misc_feature    9210. .10726
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    10827. .13819
                     /note="assembly_fragment"
     misc_feature    13920. .17308
                     /note="assembly_fragment"
     misc_feature    17409. .21780
                     /note="assembly_fragment"
     misc_feature    21881. .26334
                     /note="assembly_fragment"
     misc_feature    26435. .31916
                     /note="assembly_fragment"
     misc_feature    32017. .38192
                     /note="assembly_fragment"
     misc_feature    38293. .43903
                     /note="assembly_fragment"
     misc_feature    44004. .48720
                     /note="assembly_fragment"
     misc_feature    48821. .54929
                     /note="assembly_fragment"
     misc_feature    55030. .61995
                     /note="assembly_fragment"
     misc_feature    62096. .68368
                     /note="assembly_fragment"
     misc_feature    68469. .73768
                     /note="assembly_fragment"
     misc_feature    73869. .81655
                     /note="assembly_fragment"
     misc_feature    81756. .93898
                     /note="assembly_fragment"
     misc_feature    93999. .103949
                     /note="assembly_fragment"
     misc_feature    104050. .115686
                     /note="assembly_fragment"
     misc_feature    115787. .128950
                     /note="assembly_fragment"
     misc_feature    129051. .145870
                     /note="assembly_fragment"
     misc_feature    145971. .158947
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    49312 a  29483 c  29665 g  48081 t   2406 others
ORIGIN


alignment_scores:
            Quality:   67.00              Length:     16
              Ratio:  4.467                Gaps:      0
 Percent Similarity:  93.750   Percent Identity: 75.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC012220/rev   ..

 Align seg 1/1  to reverse of: AC012220  from: 1  to: 158947

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeu 19
         |||   ||||||||||||||||||:::::|||||||||:::||||||
  130226 CCAACTTGGCTCCTCCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTG 130179

seq_name: gb_htg2:AC009762

seq_documentation_block:
LOCUS       AC009762   179956 bp    DNA             HTG       26-FEB-2001
DEFINITION  Homo sapiens clone RP11-114I16, *** SEQUENCING IN PROGRESS ***, 4
            ordered pieces.
ACCESSION   AC009762
VERSION     AC009762.11  GI:13123283
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179956)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-114I16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179956)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
```

```
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 25, 2001 this sequence version replaced gi:12229479.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1931
                Center clone name: 114_I_16
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    17750: contig of 17750 bp in length
            *   17751 17850: gap of      100 bp
            *   17851    98658: contig of 80808 bp in length
            *   98659 98758: gap of      100 bp
            *   98759   125901: contig of 27143 bp in length
            *  125902 126001: gap of      100 bp
            *  126002   179956: contig of 53955 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179956
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-114I16"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    55479 a  35023 c  33801 g  55353 t    300 others
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     16
              Ratio:  4.467               Gaps:      0
 Percent Similarity:  93.750   Percent Identity: 75.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AC009762/rev   ..

 Align seg 1/1  to reverse of: AC009762  from: 1  to: 179956

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeu 19
         |||   |||||||||||||||||:::::|||||||||:::|||||||
   25978 CCAACTTGGCTCCTCCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTG 25931

seq_name: gb_pr9:HSSDHBS5

seq_documentation_block:
LOCUS       HSSDHBS5      649 bp    DNA             PRI       08-NOV-1995
DEFINITION  Human succinate dehydrogenase iron-protein subunit (sdhB) gene,
            exon 5.
ACCESSION   U17883
VERSION     U17883.1  GI:773295
KEYWORDS    .
SEGMENT     5 of 8
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Au,H.C., Ream-Robinson,D., Bellew,L.A., Broomfield,P.L.,
            Saghbini,M. and Scheffler,I.E.
  TITLE     Structural organization of the gene encoding the human iron-sulfur
            subunit of succinate dehydrogenase
  JOURNAL   Gene 159 (2), 249-253 (1995)
  MEDLINE   95347607
REFERENCE   2  (bases 1 to 649)
  AUTHORS   Scheffler,I.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1994) Immo E. Scheffler, Biology, University of
            California, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="lambdaDRR-A3, lambdaDRR-6A, lambdaDRR-9,
                     lambdaDRR-11, and lambdaDRR-15"
                     /clone_lib="lambda DASH and lambda FIX II human genomic
                     libraries (Stratagene)"
     intron          <1. .268
                     /gene="sdhB"
                     /number=4
     exon            269. .385
                     /gene="sdhB"
                     /number=5
     intron          386. .>649
                     /gene="sdhB"
                     /number=5
BASE COUNT      169 a    124 c    174 g    182 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     18
              Ratio:  4.125               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 66.667

alignment_block:
 US-09-215-435-179_COPY_1_23 x HSSDHBS5   ..

 Align seg 1/1  to: HSSDHBS5  from: 1  to: 649

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
         |||   |||:::|||||||:::::|||||||||:::||||||||||||
     217 TGGAATCTGATCCTTTTCTTCTTCTTCTTCTTCTTCCTCTTAACCAC 266

      22 gGly 23
         |||
     267 AGGA 270
```

```
OM of: US-09-215-435-179_COPY_1_23 to: N_Geneseq_0601:*   out_format : pfs

Date: Nov 15, 2001  4:28 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
 -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465 +  119.00  269.49  4.9e-07   421  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614 +  119.00  267.31  6.4e-07   542  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408 +  119.00  265.36  8.3e-07   680  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF44847 -   62.50  140.73  7.22      474  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA81713 -   62.00  136.31  12.73     695  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ67317 +   62.00  131.16  24.65     1264
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC35324 -   61.00  133.92  17.28     705  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80319 -   60.50  125.52  50.80     1641
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC29461 -   60.00  141.31  6.70      230  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAX14560 -   59.50  135.43  14.24     399  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74446 -   59.00  122.67  73.16     1540
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74606 +   59.00  117.47  142.55    2817
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT91972 +   59.00  112.06  285.51    5283
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV90279 -   58.00  134.41  16.23     303  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT05868 -   58.00  113.59  234.42    3399
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA64139 -   58.00   90.44  4.6e+03   50000
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA10225 -   58.00   83.91  1.1e+04   106746
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78111 -   57.00  130.60  26.48     363  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1987.DAT:AAN70775 -   57.00  122.18  77.97     965  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX01422 +   57.00  116.41  163.44    1886
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT05627 -   57.00  111.70  299.05    3259
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ60606 -   57.00   89.34  5.3e+03   43676
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ86967 -   57.00   78.03  2.2e+04   162450
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA43504 +   56.00  126.37  45.53     456  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC51928 -   56.00  125.31  52.19     516  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC42242 +   56.00  119.51  109.70    1011
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC47111 +   56.00  117.26  146.53    1314
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC50699 +   56.00  117.26  146.53    1314
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC59427 -   56.00  115.62  180.88    1590
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA09381 +   56.00  114.54  207.57    1801
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ92310 +   56.00  114.28  214.59    1856
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV15234 +   56.00  114.28  214.59    1856
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC38547 -   56.00  113.83  227.39    1956
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT65102 +   56.00  110.79  335.82    2784
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74460 +   56.00   99.44  1.4e+03   10401
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA95944 +   56.00   98.34  1.7e+03   11820
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA34791 +   56.00   77.17  2.5e+04   138169
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF20913 +   56.00   76.95  2.6e+04   141589
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21127 +   56.00   76.95  2.6e+04   141589
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21152 +   56.00   76.95  2.6e+04   141589
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA35005 +   56.00   76.95  2.6e+04   141
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA35030 +   56.00   76.95  2.6e+04   141
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21442 +   56.00   76.63  2.7e+04   146
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21437 +   56.00   73.59  3.9e+04   209
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV89689 +   55.50  128.18  36.08     324

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465

seq_documentation_block:
ID   AAX51465 standard; cDNA; 421 BP.
XX
AC   AAX51465;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:44.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12687.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 177; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 421 BP; 96 A; 106 C; 110 G; 102 T; 7 other;


alignment_scores:
           Quality:  119.00            Length:     23
             Ratio:  5.174               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
US-09-215-435-179_COPY_1_23 x AAX51465   ..

Align seg 1/1  to: AAX51465  from: 1  to: 421

      1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     62 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 111

     17 uPheLeuLeuThrArgGly 23
        |||||||||||||||||||
    112 CTTCCTCCTCACCAGGGGC 130

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614

seq_documentation_block:
ID   AAX97614 standard; DNA; 542 BP.
XX
AC   AAX97614;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 78.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY35930.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 212; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 542 BP; 132 A; 146 C; 126 G; 138 T; 0 other;


alignment_scores:
            Quality:  119.00            Length:     23
              Ratio:  5.174               Gaps:      0
```

```
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-215-435-179_COPY_1_23 x AAX97614   ..

Align seg 1/1  to: AAX97614  from: 1  to: 542

      1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

     17 uPheLeuLeuThrArgGly 23
        |||||||||||||||||||
     66 CTTCCTCCTCACCAGGGGC 84

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408

seq_documentation_block:
ID   AAA26408 standard; cDNA; 680 BP.
XX
AC   AAA26408;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 63 SEQ ID NO:73.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW   osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW   antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW   immune disease; inflammation; blood disorder; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006698-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-US17130.
XX
PR   30-JUL-1998;   98US-0094657.
PR   05-AUG-1998;   98US-0095486.
PR   06-AUG-1998;   98US-0095454.
PR   06-AUG-1998;   98US-0095455.
PR   12-AUG-1998;   98US-0096319.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Rosen CA,  Ruben SM,  Duan R,  Moore PA,  Shi Y;
PI   Lafleur D,  Wei Y,  Ni J,  Florence KA,  Young PE,  Brewer LA;
PI   Soppet DR,  Endress GA,  Ebner R,  Olsen HS,  Mucenski M;
XX
DR   WPI; 2000-195282/17.
DR   P-PSDB; AAY91513.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 416; 634pp; English.
XX
CC   The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC   human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC   antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC   antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC   corresponding secreted proteins are useful for preventing, treating or
CC   ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC   pathological conditions can be diagnosed by determining the amount of the
CC   proteins in a sample or by determining the presence of mutations in the
CC   polynucleotides. Specific uses are described for each of the
```

```
CC   polynucleotides, based on which tissues they are most highly expressed
CC   in, and include developing products for the diagnosis or treatment of
CC   cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, inflammation,
CC   allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC   osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC   transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC   cardiovascular disorders, reproductive disorders, gastrointestinal
CC   disorders, respiratory disorders and metabolic disorders. The proteins
CC   or polynucleotides can also be used as food additives or preservatives.
CC   The proteins are also useful for identifying their binding partners.
CC   AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 680 BP; 163 A; 185 C; 167 G; 156 T; 9 other;


alignment_scores:
            Quality:  119.00            Length:     23
              Ratio:  5.174              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAA26408   ..

 Align seg 1/1  to: AAA26408  from: 1  to: 680

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
     183 CTTCCTCCTCACCAGGGGC 201

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF44847

seq_documentation_block:
ID   AAF44847 standard; cDNA; 474 BP.
XX
AC   AAF44847;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human breast cancer related protein coding sequence SEQ ID NO: 3.
XX
KW   Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078960-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17536.
XX
PR   23-JUN-1999;   99US-0140903.
PR   12-OCT-1999;   99US-0158980.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Mitcham JL;
XX
DR   WPI; 2001-041426/05.
XX
PT   New polynucleotides encoding breast tumor specific proteins, useful for
PT   prevention, treatment and diagnosis of breast cancer -
XX
PS   Claim 5; Page 119; 165pp; English.
XX
CC   The present invention provides the coding sequences for a number of
CC   breast cancer related proteins. These can be used in vaccinations
CC   against, diagnosis of and treatment of cancer, particularly breast
CC   cancer.
XX
SQ   Sequence 474 BP; 191 A; 64 C; 123 G; 96 T; 0 other;


alignment_scores:
            Quality:   62.50            Length:     19
              Ratio:  3.676              Gaps:      1
 Percent Similarity:  89.474   Percent Identity: 78.947

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAF44847/rev   ..

 Align seg 1/1  to reverse of: AAF44847  from: 1  to: 474

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         :::||||||      ||||||||||||||||||:::||||||||||||||||
     100 CTTCTTCCT...CTTCTTCTTCTTCTCTTCTTGTTCTTCTTCTTCCTCTT 54

      18 eLeuLeu 20
         |||||||
      53 CCTCTTG 47

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA81713

seq_documentation_block:
ID   AAA81713 standard; DNA; 695 BP.
XX
AC   AAA81713;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_260 SEQ ID NO:260.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 1548; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
```

```
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 695 BP; 176 A; 188 C; 163 G; 168 T; 0 other;


alignment_scores:
            Quality:   62.00              Length:     18
              Ratio:  4.429                Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAA81713/rev   ..

 Align seg 1/1  to reverse of: AAA81713  from: 1  to: 695

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         :::|||   |||   :::|||   ||||||   ||||||||||||||||||
     224 GTCCCTTTGTGGCCTGTTTTAGCTTTTCTTTCTTTTTTTTTTCTTTTTCT 175

      19 uLeu 20
         ||||
     174 TTTA 171

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ67317

 seq_documentation_block:
 ID   AAQ67317 standard; DNA; 1264 BP.
 XX
 AC   AAQ67317;
 XX
 DT   27-MAR-1995  (first entry)
 XX
 DE   Sprague-Dawley rat Platelet Factor 4 promoter.
 XX
 KW   T antigen; early region; SV40; simian virus; transgenic; mammal;
 KW   megakaryocyte promoter; thrombocytopenia; decreased platelet count;
 KW   platelet precursor cell; PF4; platelet factor 4; screening;
 KW   treatment; prevention; Sprague-Dawley; Fisher; tsA58; ss.
 XX
 OS   Rattus rattus.
 XX
 PN   EP607925-A.
 XX
 PD   27-JUL-1994.
 XX
 PF   18-JAN-1994;   94EP-0100660.
 XX
 PR   19-JAN-1993;   93US-0006082.
 XX
 PA   (AMGE-) AMGEN INC.
 XX
 PI   Bosselman RA,  Hunt P,  Robinson MO;
 XX
 DR   WPI; 1994-272745/34.
 XX
 PT   Transgenic models pref. mice - used as model for thrombocytopenia
 PT   and/or megakaryocyte leukaemia
 XX
 PS   Claim 4; Page 11-12; 21pp; English.
 XX
 CC   AAQ67317 is the rat platelet factor 4 (PF4) promoter from Sprague-
 CC   Dawley strain. It was discovered that transgenic mice contg. a
 CC   nucleic acid construct encoding the SV40 early region tsA58 mutant
 CC   regulated by this promoter unexpectedly develop varying levels of
 CC   thrombocytopenia and/or megakaryocyte leukaemia. These mice present
 CC   a novel and useful system for screening cpds. useful for treating
 CC   and/or preventing these diseases.
 XX
 SQ   Sequence 1264 BP; 324 A; 283 C; 276 G; 381 T; 0 other;


 alignment_scores:
             Quality:   62.00              Length:     17
               Ratio:  3.875                Gaps:      0
  Percent Similarity:  94.118   Percent Identity: 76.471

 alignment_block:
  US-09-215-435-179_COPY_1_23 x AAQ67317   ..

  Align seg 1/1  to: AAQ67317  from: 1  to: 1264

        4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
          ||||||   ||||||||||||:::|||||||||||||||:::|||:::||
      987 CCTCAGCTGCTGCTTCTTCTTCTTCTTTTTTTTTTTTTTTTTTCCT 1036

       20 u 20
          |
     1037 G 1037

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC35324

 seq_documentation_block:
 ID   AAC35324 standard; DNA; 705 BP.
 XX
 AC   AAC35324;
 XX
 DT   17-OCT-2000  (first entry)
 XX
 DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 9777.
 XX
 KW   Hybridisation assay; genetic mapping; gene expression control;
 KW   protein identification; signal transduction pathway;
 KW   metabolic pathway; promoter; termination sequence; ss.
 XX
 OS   Arabidopsis thaliana.
 XX
 PN   EP1033405-A2.
 XX
 PD   06-SEP-2000.
 XX
 PF   25-FEB-2000; 2000EP-0301439.
 XX
 PR   25-FEB-1999;   99US-0121825.
 PR   05-MAR-1999;   99US-0123180.
 PR   09-MAR-1999;   99US-0123548.
 PR   23-MAR-1999;   99US-0125788.
 PR   25-MAR-1999;   99US-0126264.
 PR   29-MAR-1999;   99US-0126785.
 PR   01-APR-1999;   99US-0127462.
 PR   06-APR-1999;   99US-0128234.
 PR   08-APR-1999;   99US-0128714.
 PR   16-APR-1999;   99US-0129845.
 PR   19-APR-1999;   99US-0130077.
 PR   21-APR-1999;   99US-0130449.
 PR   23-APR-1999;   99US-0130510.
 PR   23-APR-1999;   99US-0130891.
 PR   28-APR-1999;   99US-0131449.
 PR   30-APR-1999;   99US-0132048.
 PR   30-APR-1999;   99US-0132407.
 PR   04-MAY-1999;   99US-0132484.
 PR   05-MAY-1999;   99US-0132485.
```

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:   61.00              Length:     17
              Ratio:  3.812                Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 70.588

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAC35324/rev   ..

 Align seg 1/1  to reverse of: AAC35324  from: 1  to: 705

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         |||||||:::   ||||||||||||||||||:::|||||||||:::|||::
      81 TTGCCACGACATCTTCTTCTTCTTTTCTTATTTTTTTTCTTCTTCTTTGT 32

      19 u 19
         :
      31 T 31

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80319

seq_documentation_block:
ID   AAQ80319 standard; DNA; 1641 BP.
XX
AC   AAQ80319;
XX
DT   22-AUG-1995  (first entry)
XX
DE   Human 38kDa FK-506 binding protein DNA.
XX
KW   Human 38kDA FK-506 binding protein; allograft rejection;
KW   T-cell activation prevention; immunosuppressants; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             141..1208
FT                   /*tag=  a
XX
PN   GB2279349-A.
XX
PD   04-JAN-1995.
XX
PF   15-JUN-1994;   94GB-0012000.
XX
PR   25-JUN-1993;   93US-0083944.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Lam E,  Wiederrecht GJ;
XX
DR   WPI; 1995-024909/04.
DR   P-PSDB; AAR64226.
XX
PT   New human 38 kD FK-506 binding protein and DNA encoding it -
PT   useful for identifying cpds. able to block allo:graft rejection
PT   by preventing T-cell activation
XX
PS   Claim 2; Pages 33-34; 44pp; English.
XX
CC   AAQ80319 encodes AAR64226 the human 38kDa FK-506 binding protein. The
CC   protein can be used to identify cpds. with immunosuppressive
CC   activity, able to block allograft rejection by preventing T-cell
CC   activation. It can also be used, opt. when immobilised, to
CC   identify, quantify and purify FK-506 and FK-506 like cpds.
XX
SQ   Sequence 1641 BP; 324 A; 550 C; 507 G; 260 T; 0 other;


alignment_scores:
            Quality:   60.50              Length:     24
              Ratio:  3.184                Gaps:      1
 Percent Similarity:  79.167   Percent Identity: 58.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAQ80319/rev   ..

 Align seg 1/1  to reverse of: AAQ80319  from: 1  to: 1641

       2 MetLeuProGlnTrpLeuLeu.........LeuLeuPheLeuLeuPhePh 15
         :::|||   :::|||   |||         :::||||||||||||||||::
     140 GTCCTCCAGCGGTGGCAGCTCACTCAGGTCATCCTCTTCCTCCTCTTCCT 91

      15 ePheLeuPheLeuLeuThrArg 22
         ::::||||||||||||||||||
      90 CCTCCTCTTCCTCCTCACCCGC 69

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC29461

seq_documentation_block:
ID   AAC29461 standard; cDNA; 230 BP.
XX
AC   AAC29461;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 33536.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 33536; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
```

```
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 230 BP; 97 A; 32 C; 32 G; 69 T; 0 other;


alignment_scores:
           Quality:   60.00             Length:     15
             Ratio:  4.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAC29461/rev   ..

 Align seg 1/1  to reverse of: AAC29461  from: 1  to: 230

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPheLeuLeu 20
         |||:::||||||:::||||||:::|||:::|||:::|||||||||
     167 TGGATATTGTTAATGTTTTTAGTTTTTATTTTCTTCTTTCTACTC 123

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAX14560

seq_documentation_block:
ID   AAX14560 standard; DNA; 399 BP.
XX
AC   AAX14560;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1544 gene.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             47..382
FT                   /*tag= a
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; AAW98841.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 1883-1884; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 399 BP; 177 A; 52 C; 82 G; 88 T; 0 other;


alignment_scores:
           Quality:   59.50             Length:     15
             Ratio:  4.250               Gaps:      1
 Percent Similarity:  93.333   Percent Identity: 86.667

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAX14560/rev   ..

 Align seg 1/1  to reverse of: AAX14560  from: 1  to: 399

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhe...PheLeuPheLeu 19
         ||||||||||||||||||:::|||||||||   ||||||||||||
     166 TGGCTTCTGCTCTTGTTTGTCTTGTTTTTCTTGTTTCTGTTCCTC 122

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74446

seq_documentation_block:
ID   AAV74446 standard; DNA; 1540 BP.
XX
AC   AAV74446;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #135.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1381..1440
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 721-722; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
```

```
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 1540 BP; 437 A; 225 C; 299 G; 517 T; 62 other;


alignment_scores:
            Quality:   59.00              Length:     17
              Ratio:  3.688                Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 58.824

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAV74446/rev   ..

 Align seg 1/1  to reverse of: AAV74446  from: 1  to: 1540

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         |||:::|||||||:::|||||||||   |||||||||:::|||:::::::
      61 CCAAAATGGCTTGTATTGCTTTTTTGTCTTTTCTTTCTTTTGTTGTTCAT 12

       20 u 20
          :
       11 G 11

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74606

seq_documentation_block:
ID   AAV74606 standard; DNA; 2817 BP.
XX
AC   AAV74606;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #295.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    601..660
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    2401..2460
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 1120-1122; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 2817 BP; 912 A; 330 C; 446 G; 999 T; 130 other;


alignment_scores:
            Quality:   59.00              Length:     18
              Ratio:  3.688                Gaps:      1
 Percent Similarity:  88.389   Percent Identity: 66.667

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAV74606   ..

 Align seg 1/1  to: AAV74606  from: 1  to: 2817

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhe......PheLeuPheLeuLe 20
         |||:::::::|||:::|||||||||||||||      ||||||||||||::
    1168 TGGTTCTTCCTAATCTTTTTACTATTCTTTGGCTTGTTTTTATTCTTAAT 1217

      20 uThr 21
         :|||
    1218 CACA 1221

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT91972

seq_documentation_block:
ID   AAT91972 standard; DNA; 5283 BP.
XX
AC   AAT91972;
XX
DT   13-FEB-1998  (first entry)
XX
DE   DNA encoding rat thrombomodulin.
XX
KW   rat; thrombomodulin; thrombin; activation; protein C;
KW   recombinant production; ds.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1386..3119
```

```
FT                   /*tag=  a
FT                   /product= thrombomodulin
XX
PN   JP09268200-A.
XX
PD   14-OCT-1997.
XX
PF   01-APR-1996;   96JP-0078494.
XX
PR   01-APR-1996;   96JP-0078494.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 1997-554712/51.
DR   P-PSDB; AAW30845.
XX
PT   Recombinant production of rat thrombomodulin - which promotes
PT   thrombin activation of protein C
XX
PS   Disclosure; Pages 22-25; 28pp; Japanese.
XX
CC   This DNA encodes a rat thrombomodulin, which upon binding thrombin,
CC   promotes thrombin activation of protein C. Host cells transformed with
CC   the DNA can be used for the large scale recombinant production of the
CC   peptide, or rat thrombomodulin.
XX
SQ   Sequence 5283 BP; 1244 A; 1324 C; 1330 G; 1385 T; 0 other;


alignment_scores:
            Quality:   59.00              Length:     17
              Ratio:  3.933                Gaps:      0
 Percent Similarity:  88.235   Percent Identity: 64.706

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAT91972   ..

 Align seg 1/1  to: AAT91972  from: 1  to: 5283

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPheLeuLeuThrAr 22
         |||:::|||   |||:::||||||||||||||||||||   :::|||::
    3874 TGGGTTTTATGATTGTTGCTGTTGTTTTTTTTTTTATTTAGAATGACAAA 3923

      22 g 22
         :
    3924 A 3924

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV90279

seq_documentation_block:
ID   AAV90279 standard; cDNA; 303 BP.
XX
AC   AAV90279;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone D1387.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 489; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 303 BP; 121 A; 21 C; 124 G; 37 T; 0 other;


alignment_scores:
            Quality:   58.00              Length:     15
              Ratio:  4.462                Gaps:      0
 Percent Similarity:  86.667   Percent Identity: 73.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x AAV90279/rev   ..

 Align seg 1/1  to reverse of: AAV90279  from: 1  to: 303

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPheLeuPhe 18
         |||   |||   ||||||||||||||||||:::::|||||||||
     216 CCAACCTGGTGCCTTTTGTTATTTCTCCTCCTACTCTTCCTCTTC 172

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT05868

seq_documentation_block:
ID   AAT05868 standard; DNA; 3399 BP.
XX
AC   AAT05868;
XX
DT   14-AUG-1996  (first entry)
XX
DE   Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
KW   Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW   immunisation; vaccination; ss.
XX
OS   Chicken leucocytozoan.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3399
FT                   /*tag=  a
FT   misc_feature    1150..3218
FT                   /*tag=  b
FT                   /note= "fragment referred to in the claims, for
FT                           use as insert in a recombinant vaccine
FT                           against chicken leucocytozoan disease"
XX
PN   JP07284392-A.
XX
PD   31-OCT-1995.
XX
PF   19-APR-1994;   94JP-0080643.
```

XX
PR 19-APR-1994; 94JP-0080643.
XX
PA (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA (KITA) KITASATO KENKYUSHO SH.
XX
DR WPI; 1996-006311/01.
DR P-PSDB; AAR97866.
XX
PT Chicken leucocytozoan immunogenic protein - used in a recombinant
PT vaccine against chicken leucocytozoan disease
XX
PS Claim 6; Page 6-9; 35pp; Japanese.
XX
CC AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC or a fragment of it can be used in a recombinant vaccine to immunise
CC against chicken leucocytozoan disease. The DNA is used in a vector
CC and operatively linked to an expression regulatory sequence as in
CC standard practice.
XX
SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;

alignment_scores:
Quality: 58.00 Length: 20
Ratio: 3.222 Gaps: 1
Percent Similarity: 90.000 Percent Identity: 65.000

alignment_block:
US-09-215-435-179_COPY_1_23 x AAT05868/rev ..

Align seg 1/1 to reverse of: AAT05868 from: 1 to: 3399

```
       1 MetMetLeuProGlnTrpLeuLeuLeuLeu.PheLeuLeuPhePhePheL 17
         ::::::||||||   :::||||||||||||| ||||||:::|||||||||:
    2869 CTTCTTCTTCCTTATTTTCTTCTTCTTCTTCCTCATTTTCTTCTTCT 2820

      17 euPheLeu 19
         ::|||||||
    2819 TCTTCCTC 2812
```

```
OM of: US-09-215-435-179_COPY_1_23 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  4:11 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-006-082A-4 +   62.00  135.02  7.18      1160  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-006-082A-1 +   62.00  134.28  7.90      1264  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-03747-1 -   57.00  114.51  99.73     3259
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-356-952-12 -   57.00   92.15  1.8e+03   43676  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-464-523B-12 +   56.00  117.03  72.12     1856  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-420-443-5 +   55.00  141.47  3.14       83  !  Se
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-446-504-63 -   55.00  116.91  73.23     1437  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-257-073-4 -   55.00  113.16  118.58    2223  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-115-446-1 -   55.00  112.07  136.26    2521  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-115-446-5 -   55.00  112.05  136.55    2526  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-446-504-83 -   55.00  109.07  200.36    3574  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-431-080-27 -   55.00  106.89  264.72    4599  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-938-534-27 -   55.00  106.89  264.72    4599  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-476-866-20 +   55.00  100.14  629.83    10079  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-009-913-287 -   54.00  138.85  4.40       86  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-342-461-5 +   54.00  123.31  32.23      522  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-081-686-1 +   54.00  114.47  100.23     1458  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-906-871-11 -   54.00  113.11  119.22     1706  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-342-461-1 +   54.00  111.92  138.90     1959  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-081-686-3 +   54.00  110.62  164.17     2279  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-131-648-4 -   54.00  110.58  165.05     2290  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-604-333-3 +   54.00  108.75  208.54     2830  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-110-618-3 +   54.00  108.75  208.54     2830  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-506-340A-2 -   54.00  108.67  210.90     2859  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-321-670-1 -   54.00  107.58  242.49     3244  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-906-871-15 -   54.00   93.15  1.5e+03   17327  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-269-040-1 +   53.50  102.03  494.26     5401  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-269-040-7 +   53.50  102.02  494.87     5407  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-370-975B-6 +   53.50   90.62  2.1e+03   20303  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-370-975B-1 +   53.50   88.24  2.9e+03   26764  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-488-470A-11 +   53.00  134.02  8.16       115  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-946-239-3 +   53.00  134.02  8.16       115  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-484-505A-11 +   53.00  134.02  8.16       115  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-151-467-3 +   53.00  134.02  8.16       115  !  S
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-036-599-7 +   53.00  134.02  8.16       115  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-078-403A-11 +   53.00  134.02  8.16       115  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-07815-3 +   53.00  134.02  8.16      115
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-594-031-165 -   53.00  129.08  15.37      204  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-170-294-1 +   53.00  115.80  84.50      954  !  S
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-664-855-1 +   53.00  115.80  84.50      954  !  S
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-049-289-1 +   53.00  115.80  84.50      954  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-328-111-333 -   52.00  116.78  74.50      650
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-06418-2 -   52.00  112.62  126.94     10
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-464-523B-2 +   52.00  108.41  217.86     1717
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-741-134-1 -   52.00  106.06  294.41     2255

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-006-082A-4

seq_documentation_block:
; Sequence 4, Application US/08006082A
; Patent No. 5489743
;  GENERAL INFORMATION:
;    APPLICANT:  ROBINSON ET AL.
;    TITLE OF INVENTION:  TRANSGENIC MAMMALS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  Amgen Center
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in., DS, 2.0 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh OS 7.0.
;      SOFTWARE:  Microsoft Word Version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,082A
;      FILING DATE:  19-JAN-1993
;      CLASSIFICATION:  800
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1160 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single stranded
;      TOPOLOGY:  Linear
US-08-006-082A-4


alignment_scores:
            Quality:   62.00              Length:     17
              Ratio:  3.875                Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 76.471

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-006-082A-4   ..

 Align seg 1/1  to: US-08-006-082A-4  from: 1  to: 1160

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         ||||||   ||||||||||||:::|||||||||||||||||:::|||:::||
     921 CCTCAGCTGCTGCTTCTTCTTCTTCTTTTTTTTTTTTTTTTTTCCT 970

      20 u 20
         |
     971 G 971

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-006-082A-1

seq_documentation_block:
; Sequence 1, Application US/08006082A
; Patent No. 5489743
;  GENERAL INFORMATION:
;    APPLICANT:  ROBINSON ET AL.
;    TITLE OF INVENTION:  TRANSGENIC MAMMALS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  Amgen Center
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
```

```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  91320-1789
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 in., DS, 2.0 MB
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  Macintosh OS 7.0.
;       SOFTWARE:  Microsoft Word Version 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/006,082A
;       FILING DATE:  19-JAN-1993
;       CLASSIFICATION:  800
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1264 base pairs
;       TYPE:  Nucleic acid
;       STRANDEDNESS:  Single stranded
;       TOPOLOGY:  Linear
US-08-006-082A-1


alignment_scores:
            Quality:   62.00              Length:      17
              Ratio:  3.875                Gaps:       0
 Percent Similarity:  94.118   Percent Identity: 76.471

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-006-082A-1    ..

 Align seg 1/1  to: US-08-006-082A-1  from: 1  to: 1264

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         ||||||   ||||||||||||:::|||||||||||||||:::|||:::||
     987 CCTCAGCTGCTGCTTCTTCTTCTTCTTTTTTTTTTTTTTTTTTCCT 1036

      20 u 20
         |
    1037 G 1037 .

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-03747-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9503747
;  GENERAL INFORMATION:
;    APPLICANT:  LA JOLLA CANCER RESEARCH FOUNDATION
;    TITLE OF INVENTION:  Brevican, A Glial Cell Proteoglycan
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03747
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Imbra, Richard J.
;      REGISTRATION NUMBER:  37,643
;      REFERENCE/DOCKET NUMBER:  FP-LJ 1453
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3259 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  112..2848
PCT-US95-03747-1


alignment_scores:
            Quality:   57.00              Length:      17
              Ratio:  3.800                Gaps:       0
 Percent Similarity:  88.235   Percent Identity: 70.588

alignment_block:
 US-09-215-435-179_COPY_1_23 x PCT-US95-03747-1/rev    ..

 Align seg 1/1  to reverse of: PCT-US95-03747-1  from: 1  to: 3259

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         ||||||   |||:::|||   :::||||||||||||||||||||||:::||
    1551 CCACAGGGCCTCATCCTCCACCTCCTCCTCTTCTTCTTCCTCTTCTTTCT 1502

      20 u 20
         |
    1501 C 1501

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-356-952-12

seq_documentation_block:
; Sequence 12, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 43676
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12


alignment_scores:
            Quality:   57.00              Length:      15
              Ratio:  3.800                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 80.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-356-952-12/rev    ..

 Align seg 1/1  to reverse of: US-09-356-952-12  from: 1  to: 43676

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         :::|||:::||||||||||||||||||:::|||||||||||||||
   20963 TTCTTGTTCCTCCTCTTCCTCCTCTTCCTCTTCCTTTTCCTCCTC 20919

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-464-523B-12

seq_documentation_block:
; Sequence 12, Application US/08464523B
; Patent No. 5723761
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Toni A. Voelker
;    APPLICANT:  Ling Yuan
;    APPLICANT:  Jean Kridl
;    APPLICANT:  Deborah Hawkins
;    APPLICANT:  Aubrey Jones
;    TITLE OF INVENTION:  Plant Acyl ACP Thioesterase
;    TITLE OF INVENTION:  Sequences
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.1
;      SOFTWARE:  Microsoft Word 5.1(a)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,523B
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/13131
;      FILING DATE:  10-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/152,004
;      FILING DATE:  10-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/261,695
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna E. Scherer
;      REGISTRATION NUMBER:  34,719
;      NAME:  Carl J. Schwedler
;      REGISTRATION NUMBER:  36,924
;      REFERENCE/DOCKET NUMBER:  CGNE 100-1WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (916) 753-6313
;      TELEFAX:  (916) 753-1510
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1856 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-464-523B-12


alignment_scores:
            Quality:   56.00              Length:     15
              Ratio:  4.000                Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 73.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-464-523B-12   ..

 Align seg 1/1  to: US-08-464-523B-12  from: 1  to: 1856

       5 GlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeu 19
         :::|||||||||||||||||:::||||||:::||||||   ||||||
     144 AGATGGCTTCTACTGCTGCTACTGCTGCTTTTTTTCCAGTTTCTT 188

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-420-443-5

seq_documentation_block:
; Sequence 5, Application US/08420443
; Patent No. 5607834
;  GENERAL INFORMATION:
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  C. Bruce Bagwell
;    TITLE OF INVENTION:  NUCLEIC ACID PROBES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/420,443
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/990,298
;      FILING DATE:  19921210
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Y. Rocky Tsao
;      REGISTRATION NUMBER:  34,053
;      REFERENCE/DOCKET NUMBER:  05663/002001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  83
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  N stands for a 6-carbon
;      OTHER INFORMATION:  molecular linker.  See Fig. 7
;      OTHER INFORMATION:  of the specification
US-08-420-443-5


alignment_scores:
            Quality:   55.00              Length:     18
              Ratio:  3.235                Gaps:      0
 Percent Similarity:  94.444   Percent Identity: 50.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-420-443-5   ..

 Align seg 1/1  to: US-08-420-443-5  from: 1  to: 83

       5 GlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuTh 21
         :::||||||:::::::::|||:::   |||||||||:::|||||||||::
      16 AAGTGGTTGTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTGTTGGT 65

      21 rArg 22
         ::::
      66 GAAA 69

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-446-504-63

seq_documentation_block:
; Sequence 63, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
;  APPLICANT:  UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
```

```
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION:  DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE:  1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/446,504
;  CURRENT FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  PCT/JP98/02845
;  PRIOR FILING DATE:  1998-06-24
;  PRIOR APPLICATION NUMBER:  JP 9-187496
;  PRIOR FILING DATE:  1997-06-26
;  PRIOR APPLICATION NUMBER:  JP 9-320692
;  PRIOR FILING DATE:  1997-11-27
;  NUMBER OF SEQ ID NOS:  92
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 63
;    LENGTH: 1437
;    TYPE: DNA
;    ORGANISM: Pyrococcus furiosus
US-09-446-504-63


alignment_scores:
            Quality:   55.00              Length:     13
              Ratio:  4.231                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 76.923

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-446-504-63/rev    ..

 Align seg 1/1  to reverse of: US-09-446-504-63  from: 1  to: 1437

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPhe 18
         |||||||:::|||:::|||||||||:::|||||||||||
    1395 TGGCTTTTCCTTTTCTTTCTTCTCCTCTTCTTTCTCTTC 1357

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-257-073-4

seq_documentation_block:
; Sequence 4, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-4


alignment_scores:
            Quality:   55.00              Length:     19
              Ratio:  3.235                Gaps:      0
 Percent Similarity:  89.474   Percent Identity: 63.158

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-257-073-4/rev    ..

 Align seg 1/1  to reverse of: US-08-257-073-4  from: 1  to: 2223

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         :::::::|||      ::::::||||||||||||||||||||||:::|||||
    2161 CTTCTTCTTTTTCTTTTTTTTCTTCTTCTTTTTCTTCTTTTTCTTTTTCT 2112

      17 uPheLeu 19
         |||||||
    2111 TTTTCTT 2105

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-115-446-1

seq_documentation_block:
; Sequence 1, Application US/09115446
; Patent No. 6165719
; GENERAL INFORMATION:
;  APPLICANT: Chandy, George K.
;  APPLICANT: Gargus, Jay J.
;  APPLICANT: Gutman, George
;  APPLICANT: Fantino, Emmanuelle
;  APPLICANT: Kalman, Katarin
;  TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM
;  TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
;  TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
;  FILE REFERENCE: 07306/014001
;  CURRENT APPLICATION NUMBER: US/09/115,446
;  CURRENT FILING DATE: 1998-07-14
;  EARLIER APPLICATION NUMBER: 60/052,556
;  EARLIER FILING DATE: 1997-07-15
;  EARLIER APPLICATION NUMBER: 60/070,741
;  EARLIER FILING DATE: 1998-01-08
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2521
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (287)...(2479)
US-09-115-446-1


alignment_scores:
            Quality:   55.00              Length:     18
              Ratio:  3.235                Gaps:      0
 Percent Similarity:  94.444   Percent Identity: 61.111

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-115-446-1/rev    ..
```

```
 Align seg 1/1  to reverse of: US-09-115-446-1  from: 1  to: 2521

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPheLeuLeuThrAr 22
         ||||||||||||||||:::||||||:::::::::|||:::|||:::   ||
     412 TGGCTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTCATCCCCAG 363

      22 gGly 23
         ||||
     362 AGGA 359

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-115-446-5

seq_documentation_block:
; Sequence 5, Application US/09115446
; Patent No. 6165719
; GENERAL INFORMATION:
;  APPLICANT: Chandy, George K.
;  APPLICANT: Gargus, Jay J.
;  APPLICANT: Gutman, George
;  APPLICANT: Fantino, Emmanuelle
;  APPLICANT: Kalman, Katarin
;  TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM
;  TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
;  TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
;  FILE REFERENCE: 07306/014001
;  CURRENT APPLICATION NUMBER: US/09/115,446
;  CURRENT FILING DATE: 1998-07-14
;  EARLIER APPLICATION NUMBER: 60/052,556
;  EARLIER FILING DATE: 1997-07-15
;  EARLIER APPLICATION NUMBER: 60/070,741
;  EARLIER FILING DATE: 1998-01-08
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 2526
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-115-446-5


alignment_scores:
            Quality:   55.00             Length:     18
              Ratio:  3.235              Gaps:        0
 Percent Similarity:  94.444   Percent Identity: 61.111

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-115-446-5/rev   ..

 Align seg 1/1  to reverse of: US-09-115-446-5  from: 1  to: 2526

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPheLeuLeuThrAr 22
         ||||||||||||||||:::||||||:::::::::|||:::|||:::   ||
     417 TGGCTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTCATCCCCAG 368

      22 gGly 23
         ||||
     367 AGGA 364

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-446-504-83

seq_documentation_block:
; Sequence 83, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
;  APPLICANT:  UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION:  DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE:  1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/446,504
;  CURRENT FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  PCT/JP98/02845
;  PRIOR FILING DATE:  1998-06-24
;  PRIOR APPLICATION NUMBER:  JP 9-187496
;  PRIOR FILING DATE:  1997-06-26
;  PRIOR APPLICATION NUMBER:  JP 9-320692
;  PRIOR FILING DATE:  1997-11-27
;  NUMBER OF SEQ ID NOS:  92
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 83
;   LENGTH: 3574
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
US-09-446-504-83


alignment_scores:
            Quality:   55.00             Length:     13
              Ratio:  4.231              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 76.923

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-446-504-83/rev   ..

 Align seg 1/1  to reverse of: US-09-446-504-83  from: 1  to: 3574

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePheLeuPhe 18
         ||||||:::|||:::||||||||||:::|||||||||||
    1398 TGGCTTTTCCTTTTCTTTCTTCTCCTCTTCTTTCTCTTC 1360

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-431-080-27

seq_documentation_block:
; Sequence 27, Application US/08431080
; Patent No. 5698686
;  GENERAL INFORMATION:
;    APPLICANT:  Gottschling, Daniel E.
;    APPLICANT:  Singer, Miriam S.
;    TITLE OF INVENTION:  Telomerase Compositions and Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TEXAS
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/431,080
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SN 08/326,781
;      FILING DATE:  October 20, 1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:155/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  27:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4599 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-431-080-27


alignment_scores:
            Quality:   55.00              Length:      15
              Ratio:  3.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 73.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-431-080-27/rev    ..

 Align seg 1/1  to reverse of: US-08-431-080-27  from: 1  to: 4599

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         :::||||||||||:::||||||:::||||||||||:::|||||||||
    1524 TTTTTGTTGCTCTTCTTTCTCTTTTTCTTTTTCTTCTTTCTGCTC 1480

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-938-534-27

seq_documentation_block:
; Sequence 27, Application US/08938534
; Patent No. 5916752
;  GENERAL INFORMATION:
;    APPLICANT:  Gottschling, Daniel E.
;    APPLICANT:  Singer, Miriam S.
;    TITLE OF INVENTION:  Telomerase Compositions and Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TEXAS
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,534
;      FILING DATE:  26-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/431,080
;      FILING DATE:
;      APPLICATION NUMBER:  SN 08/326,781
;      FILING DATE:  October 20, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:155/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4599 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-938-534-27


alignment_scores:
            Quality:   55.00              Length:      15
              Ratio:  3.667                Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 73.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-938-534-27/rev    ..

 Align seg 1/1  to reverse of: US-08-938-534-27  from: 1  to: 4599

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         :::||||||||||:::||||||:::||||||||||:::|||||||||
    1524 TTTTTGTTGCTCTTCTTTCTCTTTTTCTTTTTCTTCTTTCTGCTC 1480

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-476-866-20

seq_documentation_block:
; Sequence 20, Application US/08476866
; Patent No. 5994339
;  GENERAL INFORMATION:
;    APPLICANT:  CRAPO, JAMES D.
;    APPLICANT:  FRIDOVICH, IRWIN
;    APPLICANT:  OURY, TIM
;    APPLICANT:  DAY, BRIAN J.
;    APPLICANT:  FOLZ, RODNEY J.
;    APPLICANT:  FREEMAN, BRUCE A.
;    TITLE OF INVENTION:  SUPEROXIDE DISMUTASE AND MIMETICS THEREOF
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,866
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/322,766
;      FILING DATE:  13-OCT-1994
;      APPLICATION NUMBER:  US 08/136,207
;      FILING DATE:  15-OCT-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILSON, MARY J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  1579-74
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5086..5803
US-08-476-866-20


alignment_scores:
            Quality:   55.00              Length:      17
              Ratio:  3.438                Gaps:       0
 Percent Similarity:  94.118   Percent Identity: 64.706
```

```
alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-476-866-20   ..

 Align seg 1/1  to: US-08-476-866-20  from: 1  to: 10079

       7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArgGl 23
         ||||||||||||||||:::::::|||||||||:::|||:::|||   :::||
    1992 CTCCTTCTTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACATGG 2041

      23 y 23
         |
    2042 T 2042

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-009-913-287

seq_documentation_block:
; Sequence 287, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  287:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  86 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-009-913-287


alignment_scores:
           Quality:   54.00          Length:     15
             Ratio:  3.857             Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 73.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-009-913-287/rev   ..

 Align seg 1/1  to reverse of: US-09-009-913-287  from: 1  to: 86

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         :::|||::::::|||||||||   ||||||||||||||||||||
      63 TTTCTTTTCTTTCTCTTTCTTACTTTCTTCTTTCTTTTTCTTCTT 19
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-179_COPY_1_23 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44 +  119.00  236.19  0.0008    421
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-731-872-89 +  119.00  234.53  0.0010    514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5913 +  119.00  234.53  0.0010    514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-876-997-89 +  119.00  234.53  0.0010    514
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-1374 +  119.00  234.53  0.0010    51
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-108 +  119.00  234.53  0.0010    514
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-108 +  119.00  234.53  0.0010    514
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5913 +  119.00  234.53  0.0010    51
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817 +  119.00  234.23  0.0011    533
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817 +  119.00  234.23  0.0011    53
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78 +  119.00  234.09  0.0011    542
/cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61 +  119.00  233.74  0.0011    565
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73 +  119.00  232.20  0.0014    680
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73 +  119.00  232.20  0.0014    680
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-192 +  119.00  186.18  0.5005    171
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-219 +  119.00  186.18  0.5005    171
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481-520 +   70.00  139.89  189.44    488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481A-520 +   70.00  139.89  189.44    488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481B-520 +   70.00  139.89  189.44    488
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-244-000A-84931 -   69.00  143.55  118.52     2
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-242-679-659 +   67.00  101.89  2.5e+04    233
/cgn1_7/ptodata/1/pna/US6023_COMB.seq:US-60-230-435-750 +   67.00   99.46  3.4e+04    313
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-213-172-34 -   67.00   99.08  3.6e+04   3276
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-213-173-34 -   67.00   99.08  3.6e+04   3276
/cgn1_7/ptodata/1/pna/US096D_COMB.seq:US-09-666-355A-7255 -   66.00  133.15  450.06     4
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US01-00663-969 +   66.00  132.08  515.91    491
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-864-761-931 +   66.00  132.08  515.91    491
/cgn1_7/ptodata/1/pna/US6023_COMB.seq:US-60-236-359-1002 +   66.00  132.08  515.91     49
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-197755 -   66.00  129.19  747.41
/cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-197755 -   66.00  129.19  747.41
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-505-1523 +   66.00  123.53  1.5e+03    13
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-505-1819 -   66.00  121.88  1.9e+03    16
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-245-225-592 -   66.00  117.19  3.5e+03   294
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-950-083-7308 +   66.00  103.85  1.9e+04   146
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US01-01324-3759 +   66.00  101.14  2.7e+04    20
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US01-01350-173 +   66.00  101.14  2.7e+04   202
/cgn1_7/ptodata/1/pna/US097C_COMB.seq:US-09-764-855-173 +   66.00  101.14  2.7e+04   202
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-950-083-5384 +   66.00  101.14  2.7e+04   202
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-505-191 +   66.00   95.56  5.6e+04   396
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-248-505-487 +   66.00   95.56  5.6e+04   396
/cgn1_7/ptodata/1/pna/US6024_COMB.seq:US-60-245-225-122 +   66.00   90.26  1.1e+05
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-948-941-648 +   66.00   87.27  1.6e+05   1
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-705-926-2711 +   65.00  131.70  541.64
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-310777 +   65.00  126.96  995.40
/cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-310777 +   65.00  126.96  995.40

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44

seq_documentation_block:
; Sequence 44, Application US/08905279
;  GENERAL INFORMATION:
;    APPLICANT:  Genset SA
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,279
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.011A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  421 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  62..130
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLTRG/SL
US-08-905-279-44


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-08-905-279-44   ..

 Align seg 1/1  to: US-08-905-279-44  from: 1  to: 421

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      62 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 111

      17 uPheLeuLeuThrArgGly 23
         ||||||||||||||||||
```

```
    112 CTTCCTCCTCACCAGGGGC 130

seq_name: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-731-872-89

seq_documentation_block:
; Sequence 89, Application US/09731872
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Jobert, Severin
;  TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
;  FILE REFERENCE: 78.US3.REG
;  CURRENT APPLICATION NUMBER: US/09/731,872
;  CURRENT FILING DATE:  2000-12-07
;  PRIOR APPLICATION NUMBER: US 60/169,629
;  PRIOR FILING DATE: 1999-12-08
;  PRIOR APPLICATION NUMBER: US 60/187,470
;  PRIOR FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 482
;  SOFTWARE: Patent.pm
; SEQ ID NO 89
;   LENGTH: 514
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 113..364
;   NAME/KEY: sig_peptide
;   LOCATION: 113..172
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.37180298395146
;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
US-09-731-872-89


alignment_scores:
            Quality:  119.00            Length:     23
              Ratio:  5.174               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-731-872-89    ..

 Align seg 1/1  to: US-09-731-872-89  from: 1  to: 514

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

 seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5913

 seq_documentation_block:
 ; Sequence 5913, Application US/09834366
 ; GENERAL INFORMATION:
 ;  APPLICANT: Bejanin, Stephane
 ;  APPLICANT:  Tanaka, Hiroaki
 ;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Jobert, Severin
 ;  APPLICANT:  Giordano, Jean-Yves
 ;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ;  FILE REFERENCE: 81.US2.REG
 ;  CURRENT APPLICATION NUMBER: US/09/834,366
 ;  CURRENT FILING DATE:  2001-04-13
 ;  PRIOR APPLICATION NUMBER: US 60/197,873
 ;  PRIOR FILING DATE: 2000-04-18
 ;  NUMBER OF SEQ ID NOS: 52153
 ;  SOFTWARE: Patent.pm
 ; SEQ ID NO 5913
 ;   LENGTH: 514
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 113..364
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 113..172
 ;   OTHER INFORMATION: Von Heijne matrix
 ;   OTHER INFORMATION: score 4.30000019073486
 ;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
 US-09-834-366-5913


 alignment_scores:
             Quality:  119.00            Length:     23
               Ratio:  5.174               Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-179_COPY_1_23 x US-09-834-366-5913    ..

  Align seg 1/1  to: US-09-834-366-5913  from: 1  to: 514

        1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

       17 uPheLeuLeuThrArgGly 23
          |||||||||||||||||||
      149 CTTCCTCCTCACCAGGGGC 167

 seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-876-997-89

 seq_documentation_block:
 ; Sequence 89, Application US/09876997
 ; GENERAL INFORMATION:
 ;  APPLICANT: Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Bougueleret, Lydie
 ;  APPLICANT:  Jobert, Severin
 ;  TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
 ;  FILE REFERENCE: 78.US4.CIP
 ;  CURRENT APPLICATION NUMBER: US/09/876,997
 ;  CURRENT FILING DATE:  2001-06-08
 ;  PRIOR APPLICATION NUMBER: US 09/731,872
 ;  PRIOR FILING DATE: 2000-12-07
 ;  PRIOR APPLICATION NUMBER: US 60/187,470
 ;  PRIOR FILING DATE: 2000-03-06
 ;  PRIOR APPLICATION NUMBER: US 60/169,629
 ;  PRIOR FILING DATE: 1999-12-08
 ;  NUMBER OF SEQ ID NOS: 482
 ;  SOFTWARE: Patent.pm
 ; SEQ ID NO 89
 ;   LENGTH: 514
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 113..364
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 113..172
 ;   OTHER INFORMATION: Von Heijne matrix
 ;   OTHER INFORMATION: score 4.37180298395146
 ;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
 US-09-876-997-89


 alignment_scores:
             Quality:  119.00            Length:     23
               Ratio:  5.174               Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-179_COPY_1_23 x US-09-876-997-89    ..
```

```
 Align seg 1/1  to: US-09-876-997-89  from: 1  to: 514

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

seq_name: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-1374

seq_documentation_block:
; Sequence 1374, Application US/60147499
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Jobert, S.
;  APPLICANT: Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/60/147,499
;  CURRENT FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 1374
;   LENGTH: 514
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 99..356
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 99..167
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.80000019073486
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-60-147-499-1374


alignment_scores:
            Quality:  119.00             Length:     23
              Ratio:  5.174                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-147-499-1374   ..

 Align seg 1/1  to: US-60-147-499-1374  from: 1  to: 514

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

seq_name: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-108

seq_documentation_block:
; Sequence 108, Application US/60169629
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: GENSET.071PRF
;  CURRENT APPLICATION NUMBER: US/60/169,629
;  CURRENT FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 108
;   LENGTH: 514
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 113..364
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 113..172
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.3
;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 478..483
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 500..514
US-60-169-629-108


alignment_scores:
            Quality:  119.00             Length:     23
              Ratio:  5.174                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-169-629-108   ..

 Align seg 1/1  to: US-60-169-629-108  from: 1  to: 514

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-108

seq_documentation_block:
; Sequence 108, Application US/60187470
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: 78.US2.PRO
;  CURRENT APPLICATION NUMBER: US/60/187,470
;  CURRENT FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 108
;   LENGTH: 514
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 113..364
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 113..172
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 4.3
;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 478..483
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 500..514
```

```
US-60-187-470-108


alignment_scores:
            Quality:  119.00              Length:      23
              Ratio:  5.174                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-187-470-108    ..

 Align seg 1/1  to: US-60-187-470-108  from: 1  to: 514

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         ||||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

 seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5913

 seq_documentation_block:
 ; Sequence 5913, Application US/60197873
 ; GENERAL INFORMATION:
 ;  APPLICANT:  Bejanin, Stephane
 ;  APPLICANT:  Tanaka, Hiroaki
 ;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Jobert, Severin
 ;  APPLICANT:  Giordano, Jean-Yves
 ;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
 ;  FILE REFERENCE:  81.US1.PRO
 ;  CURRENT APPLICATION NUMBER: US/60/197,873
 ;  CURRENT FILING DATE:  2000-04-18
 ;  NUMBER OF SEQ ID NOS:  52153
 ;  SOFTWARE:  Patent.pm
 ; SEQ ID NO 5913
 ;   LENGTH: 514
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 113..364
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 113..172
 ;   OTHER INFORMATION: Von Heijne matrix
 ;   OTHER INFORMATION: score 4.30000019073486
 ;   OTHER INFORMATION: seq SLLLSLPPHQGLT/FS
 US-60-197-873-5913


 alignment_scores:
             Quality:  119.00              Length:      23
               Ratio:  5.174                Gaps:       0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-179_COPY_1_23 x US-60-197-873-5913    ..

  Align seg 1/1  to: US-60-197-873-5913  from: 1  to: 514

        1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

       17 uPheLeuLeuThrArgGly 23
          ||||||||||||||||||
      149 CTTCCTCCTCACCAGGGGC 167

  seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817

  seq_documentation_block:
```

```
; Sequence 4817, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 4817
;   LENGTH: 533
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 99..461
;   NAME/KEY: sig_peptide
;   LOCATION: 99..167
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.800000019073486
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-09-834-366-4817


alignment_scores:
            Quality:  119.00              Length:      23
              Ratio:  5.174                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-834-366-4817    ..

 Align seg 1/1  to: US-09-834-366-4817  from: 1  to: 533

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         ||||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

 seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817

 seq_documentation_block:
 ; Sequence 4817, Application US/60197873
 ; GENERAL INFORMATION:
 ;  APPLICANT:  Bejanin, Stephane
 ;  APPLICANT:  Tanaka, Hiroaki
 ;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Jobert, Severin
 ;  APPLICANT:  Giordano, Jean-Yves
 ;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
 ;  FILE REFERENCE:  81.US1.PRO
 ;  CURRENT APPLICATION NUMBER: US/60/197,873
 ;  CURRENT FILING DATE:  2000-04-18
 ;  NUMBER OF SEQ ID NOS:  52153
 ;  SOFTWARE:  Patent.pm
 ; SEQ ID NO 4817
 ;   LENGTH: 533
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: 99..461
 ;   NAME/KEY: sig_peptide
 ;   LOCATION: 99..167
```

```
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.80000019073486
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-60-197-873-4817


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-197-873-4817   ..

 Align seg 1/1  to: US-60-197-873-4817  from: 1  to: 533

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 148

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
     149 CTTCCTCCTCACCAGGGGC 167

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78

seq_documentation_block:
; Sequence 78, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 78
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 16..378
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 16..84
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.8
;   OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 502..507
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 522..542
US-09-215-435-78


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-215-435-78   ..

 Align seg 1/1  to: US-09-215-435-78  from: 1  to: 542

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
      66 CTTCCTCCTCACCAGGGGC 84

seq_name: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61

seq_documentation_block:
; Sequence 61, Application US/60096116
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  95
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/096,116
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.038PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  565 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  16..84
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLTRG/SL
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  502..507
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  529..542
;      IDENTIFICATION METHOD:  blastn2
;    FEATURE:
;      NAME/KEY:  vrt
;      LOCATION:  complement(33..74)
;      IDENTIFICATION METHOD:  fasta
;      OTHER INFORMATION:  identity 88
```

```
;       OTHER INFORMATION:  region 21722..21764
;       OTHER INFORMATION:  id AF048728
;     FEATURE:
;       NAME/KEY:  vrt
;       LOCATION:  511..545
;       IDENTIFICATION METHOD:  fasta
;       OTHER INFORMATION:  identity 83
;       OTHER INFORMATION:  region 4627..4663
;       OTHER INFORMATION:  id AF048728
US-60-096-116-61


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-096-116-61    ..

 Align seg 1/1  to: US-60-096-116-61  from: 1  to: 565

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      16 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 65

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
      66 CTTCCTCCTCACCAGGGGC 84

seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73

seq_documentation_block:
; Sequence 73, Application PC/TUS9917130
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/17130
;  CURRENT FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-17130-73


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x PCT-US99-17130-73    ..

 Align seg 1/1  to: PCT-US99-17130-73  from: 1  to: 680

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
     183 CTTCCTCCTCACCAGGGGC 201

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73

seq_documentation_block:
; Sequence 73, Application US/09489847
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031P1
;  CURRENT APPLICATION NUMBER: US/09/489,847
;  CURRENT FILING DATE: 2000-01-24
;  EARLIER APPLICATION NUMBER: PCT/US99/17130
;  EARLIER FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 376
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-489-847-73
```

```
alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-489-847-73   ..

 Align seg 1/1  to: US-09-489-847-73  from: 1  to: 680

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     133 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 182

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
     183 CTTCCTCCTCACCAGGGGC 201

seq_name: /cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-192

seq_documentation_block:
; Sequence 192, Application US/60212664
; GENERAL INFORMATION:
;  APPLICANT:  Ladunga, Steve
;  APPLICANT:  Spier, Gene
;  APPLICANT:  Greenberg, Simon
;  APPLICANT:  Rabkin, Steven
;  APPLICANT:  Wang, Yu
;  TITLE OF INVENTION:  ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL000687
;  CURRENT APPLICATION NUMBER: US/60/212,664
;  CURRENT FILING DATE:  2000-06-20
;  NUMBER OF SEQ ID NOS:  636
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 192
;   LENGTH: 171973
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(171973)
;   OTHER INFORMATION: n = A,T,C or G
US-60-212-664-192


alignment_scores:
            Quality:  119.00              Length:     23
              Ratio:  5.174                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-60-212-664-192   ..

 Align seg 1/1  to: US-60-212-664-192  from: 1  to: 171973

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   83353 ATGATGCTACCCCAATGGCTGCTGCTGCTGTTCCTTCTCTTCTTCTTTCT 83402

      17 uPheLeuLeuThrArgGly 23
         |||||||||||||||||||
   83403 CTTCCTCCTCACCAGGGGC 83421
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-179_COPY_1_23 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-192 +   60.00  119.47  19.36     9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-938-301-159 -   59.00  113.86  39.73     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-730-28 -   55.50  111.56  53.41     79
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-108-010A-11 -   55.50  106.69  99.68     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-10083 -   54.50  114.34  37.36
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-958-480-23 -   54.00  126.88  7.48      74
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-10961 -   54.00  113.13  43.66
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-789 +   54.00   83.00  2.1e+03   2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-792 +   54.00   82.89  2.1e+03   2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-768-185A-1 +   54.00   61.90  2.9e+04   32
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-2910 +   53.00  111.32  55.02
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-2909 +   53.00  110.41  61.84
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-236 +   53.00  105.12  121.97    9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-414 +   53.00  105.04  123.12    9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-730-67 -   53.00  102.51  170.38    13
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-5593 -   52.50  111.03  57.11
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2366 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2367 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2368 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2369 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2370 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2371 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2372 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2373 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2374 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2375 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2376 -   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2377 +   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2378 +   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2379 +   52.50   82.37  2.3e+03
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-750-456-160 +   52.00  109.72  67.58     4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-11118 -   52.00  109.36  70.76
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-4989 -   52.00  108.03  83.98
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3789 -   52.00  100.61  217.40
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-320 -   52.00   98.47  285.94    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4279 -   52.00   96.84  352.55
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-8224 -   52.00   96.74  357.02
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-306 +   52.00   85.58  1.5e+03   9
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-858-11 -   52.00   81.64  2.5e+03   15
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-494 -   51.00  110.44  61.63
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-13867 +   51.00  107.69  87.75
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-963-299-26 +   51.00  105.60  114.68
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-729-658B-6 +   51.00  105.25  119.88
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4345 -   51.00   98.73  276.87
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-708-205 -   51.00   94.32  487.19

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-192

seq_documentation_block:
; Sequence 192, Application US/09886055
; GENERAL INFORMATION:
;  APPLICANT: STRYER, LUBERT
;  APPLICANT:  ZOZULYA, SERGEY
;  TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
;  TITLE OF INVENTION:  BIOSENSORS OF CHEMICAL SENSANTS
;  FILE REFERENCE: 078003-0277150
;  CURRENT APPLICATION NUMBER: US/09/886,055
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: 60/213,812
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 522
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 192
;   LENGTH: 933
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-886-055-192


alignment_scores:
            Quality:   60.00              Length:      18
              Ratio:  3.750                Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 55.556

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-886-055-192   ..

 Align seg 1/1  to: US-09-886-055-192  from: 1  to: 933

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         |||:::|||   |||   ||||||||||||:::::|||::::::|||::
      61 CCTGAATGGGCACTCCCTCTCTTCCTCTTGTTATTATTTATGTATCTCAT 110

      20 uThr 21
         :|||
     111 CACC 114

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-938-301-159

seq_documentation_block:
; Sequence 159, Application US/09938301
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: Novel Nucleic Acid and Protein Homologs
;  FILE REFERENCE: 5800-155
;  CURRENT APPLICATION NUMBER: US/09/938,301
;  CURRENT FILING DATE:  2001-08-23
;  PRIOR APPLICATION NUMBER: 09/610,154
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 159
;   LENGTH: 1479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-938-301-159


alignment_scores:
            Quality:   59.00              Length:      23
              Ratio:  3.278                Gaps:       0
 Percent Similarity:  78.261   Percent Identity: 60.870
```

```
alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-938-301-159/rev   ..

 Align seg 1/1  to reverse of: US-09-938-301-159  from: 1  to: 1479

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ::::::   |||:::      ||||||||||||||||||||||||||   ||
    1084 CTCCTGCATCCCCGGCCCAAACTCCTCCTCTTCCTCCTCTTCTTCCATCT 1035

      17 uPheLeuLeuThrArgGly 23
         |:::|||||||   ||||||
    1034 CCTCCTCCTCTGACGAGGA 1016

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-730-28

seq_documentation_block:
; Sequence 28, Application US/09969730
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 90 Human Secreted Proteins
;  FILE REFERENCE: PZ013P2
;  CURRENT APPLICATION NUMBER: US/09/969,730
;  CURRENT FILING DATE:  2001-10-04
;  PRIOR APPLICATION NUMBER: 09/774,639
;  PRIOR FILING DATE: 2001-02-01
;  PRIOR APPLICATION NUMBER: 60/238,291
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 09/244,112
;  PRIOR FILING DATE: 1999-02-04
;  PRIOR APPLICATION NUMBER: PCT/US98/16235
;  PRIOR FILING DATE: 1998-08-04
;  PRIOR APPLICATION NUMBER: 60/056,371
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,732
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,366
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,364
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,370
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,367
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,365
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,731
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,557
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,563
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/055,970
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,986
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,311
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,808
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,803
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,804
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,809
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,806
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,310
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,798
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,309
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,312
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,807
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,386
;  PRIOR FILING DATE: 1997-08-05
;  NUMBER OF SEQ ID NOS: 373
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 794
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (345)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-969-730-28


alignment_scores:
            Quality:   55.50              Length:      22
              Ratio:  2.775                Gaps:       1
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-969-730-28/rev   ..

 Align seg 1/1  to reverse of: US-09-969-730-28  from: 1  to: 794

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         :::|||:::::::::|||||||:::|||   |||:::::|||||||:::||
     579 ATGCCCAGCTTCATCTTGTTATTTTTTTATTTATTATTATTTTTATACTT 530

      19 uLeu...ThrArgGly 23
         ||||   |||||||||
     529 GCTATGGACTAGGGGA 514

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-108-010A-11

seq_documentation_block:
; Sequence 11, Application US/09108010A
;  GENERAL INFORMATION:
;    APPLICANT:  ANTHONY JOHN KINNEY
;    APPLICANT:  GARY MICHAEL FADER
;    TITLE OF INVENTION:  SUPPRESSION OF SPECIFIC
;    TITLE OF INVENTION:  CLASSES OF SOYBEAN SEED
;    TITLE OF INVENTION:  PROTEIN GENES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROFSOFT WORD 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/108,010A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/019,940
;      FILING DATE:  JUNE 14,1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LYNNE M. CHRISTENBURY
;      REGISTRATION NUMBER:  30,971
;      REFERENCE/DOCKET NUMBER:  BB-1071-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-5481
```

```
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-108-010A-11


alignment_scores:
            Quality:   55.50             Length:     24
              Ratio:  2.921              Gaps:        1
 Percent Similarity:  79.167   Percent Identity: 54.167

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-108-010A-11/rev   ..

 Align seg 1/1  to reverse of: US-09-108-010A-11  from: 1  to: 1488

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePhe..... 16
         |||      |||||||||:::|||||||:::::|||:::|||
     873 CTTGCACTGTGGCTTCTCATCCTCTTCTTCTTCTTCCTCTTCCTGGGGTC 824

      17 .LeuPheLeuLeuThrArgGly 23
          |||:::|||||||:::||||||
     823 TTTGTTGCTGCTCGTCCGTGGG 802

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-10083

seq_documentation_block:
; Sequence 10083, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10083
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(426)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-10083


alignment_scores:
            Quality:   54.50             Length:     23
              Ratio:  3.028              Gaps:        1
 Percent Similarity:  78.261   Percent Identity: 56.522

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-922-340-10083/rev   ..

 Align seg 1/1  to reverse of: US-09-922-340-10083  from: 1  to: 426

       2 MetLeuProGlnTrpLeuLeu.........LeuLeuPheLeuLeuPhePh 15
         :::|||   :::|||   |||         :::||||||||||||||::
     346 GTCCTCCANCGGTGGCAGCTCANTCAGGTCATCCTCTTCCTCCTCTTCCT 297

      15 ePheLeuPheLeuLeuThr 21
         ::::|||||||||||||||
     296 CCTCCTCTTCCTCCTCACC 278

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-958-480-23

seq_documentation_block:
; Sequence 23, Application US/09958480
; GENERAL INFORMATION:
;  APPLICANT: Finney, Helen Margaret
;  APPLICANT:  Lawson, Alastair David Griffiths
;  TITLE OF INVENTION: A Method Of Altering The Properties Of a Membrane-Associated P
;  TITLE OF INVENTION:  Substitution Of The Transmembrane Domain
;  FILE REFERENCE: CARP0093
;  CURRENT APPLICATION NUMBER: US/09/958,480
;  CURRENT FILING DATE:  2001-10-11
;  PRIOR APPLICATION NUMBER: PCT/GB00/01471
;  PRIOR FILING DATE: 2000-04-17
;  PRIOR APPLICATION NUMBER: GB9908816.3
;  PRIOR FILING DATE: 1999-04-16
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 74
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: A8350
US-09-958-480-23


alignment_scores:
            Quality:   54.00             Length:     24
              Ratio:  3.176              Gaps:        1
 Percent Similarity:  70.833   Percent Identity: 58.333

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-958-480-23/rev   ..

 Align seg 1/1  to reverse of: US-09-958-480-23  from: 1  to: 74

       4 ProGlnTrpLeu.............LeuLeuLeuPheLeuLeuPhePhePh 16
         |||   ||||||             ||||||||||||   |||:::::||
      74 CCCTTGTGGCTGGGGGTCCTGCGGCTGCTGCTCTTCAAGCTGCTGCTGTT 25

      16 eLeuPheLeuLeuThrArgGly 23
         |   :::|||||||||||||||
      24 TGACCTGCTCCTGACGCGTGGA 3

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-10961

seq_documentation_block:
; Sequence 10961, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10961
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(437)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-10961
```

```
alignment_scores:
            Quality:   54.00              Length:     17
              Ratio:  3.857                Gaps:      0
 Percent Similarity:  82.353   Percent Identity: 64.706

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-922-340-10961/rev    ..

 Align seg 1/1  to reverse of: US-09-922-340-10961  from: 1  to: 437

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         |||      ||||||   ::::::||||||||||||||||:::||||||||
     275 CTGGAAGTTTGGCTCAGGTTCTTTTTTCTTCTCTTTTTCCTCCTTTTTCT 226

      19 u 19
         |
     225 C 225

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-789

seq_documentation_block:
; Sequence 789, Application US/09973278
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P2
;  CURRENT APPLICATION NUMBER: US/09/973,278
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: 60/239,899
;  PRIOR FILING DATE: 2000-10-13
;  PRIOR APPLICATION NUMBER: 09/227,357
;  PRIOR FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: PCT/US98/13684
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/051,926
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,793
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,925
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,929
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,803
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,732
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,931
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,932
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,916
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,930
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,918
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,920
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,733
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,795
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,919
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,928
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/055,722
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,723
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,948
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,949
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,953
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,950
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,947
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/056,360
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,684
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,984
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,954
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/058,785
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,664
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,660
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,661
;  PRIOR FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 947
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 789
;   LENGTH: 21358
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-973-278-789


alignment_scores:
            Quality:   54.00              Length:     13
              Ratio:  4.154                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 69.231

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-973-278-789    ..

 Align seg 1/1  to: US-09-973-278-789  from: 1  to: 21358

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPhe 18
         |||:::||||||:::|||:::||||||||||||:::|||
   17916 TGGTTCCTTTTGTTTTTTGTTCTTTTTTTTTTTTTTTTT 17954

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-792

seq_documentation_block:
; Sequence 792, Application US/09973278
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P2
;  CURRENT APPLICATION NUMBER: US/09/973,278
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: 60/239,899
;  PRIOR FILING DATE: 2000-10-13
;  PRIOR APPLICATION NUMBER: 09/227,357
;  PRIOR FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: PCT/US98/13684
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/051,926
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,793
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,925
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,929
```

```
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,803
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,732
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,931
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,932
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,916
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,930
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,918
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,920
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,733
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/052,795
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,919
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/051,928
;  PRIOR FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: 60/055,722
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,723
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,948
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,949
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,953
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,950
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,947
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/056,360
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,684
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,984
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,954
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/058,785
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,664
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,660
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,661
;  PRIOR FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 947
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 792
;   LENGTH: 21676
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-973-278-792


alignment_scores:
            Quality:   54.00            Length:     13
              Ratio:  4.154              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 69.231

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-973-278-792   ..

 Align seg 1/1  to: US-09-973-278-792  from: 1  to: 21676

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPhe 18
         |||:::|||||||:::|||:::|||||||||||||:::|||
   17916 TGGTTCCTTTTGTTTTTTGTTCTTTTTTTTTTTTTTTTT 17954

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-768-185A-1

seq_documentation_block:
; Sequence 1, Application US/09768185A
; GENERAL INFORMATION:
;  APPLICANT: Cassel, Michael et al
;  TITLE OF INVENTION: Estrogen receptor beta variants and
;  TITLE OF INVENTION:  methods of detection thereof
;  FILE REFERENCE: CL000280
;  CURRENT APPLICATION NUMBER: US/09/768,185A
;  CURRENT FILING DATE:  2001-01-24
;  PRIOR APPLICATION NUMBER: 09768185
;  PRIOR FILING DATE: 2001-01-24
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 325791
;   TYPE: DNA
;   ORGANISM: HUMAN
US-09-768-185A-1


alignment_scores:
            Quality:   54.00            Length:     16
              Ratio:  3.600              Gaps:      0
 Percent Similarity:  93.750   Percent Identity: 62.500

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-768-185A-1   ..

 Align seg 1/1  to: US-09-768-185A-1  from: 1  to: 325791

       7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrArg 22
         :::|||||||:::|||||||:::|||||||||:::|||:::|||   |||
   72589 ATTTTATTAATTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGG 72636

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-2910

seq_documentation_block:
; Sequence 2910, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2910
;   LENGTH: 424
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(424)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-2910


alignment_scores:
            Quality:   53.00            Length:     16
```

```
             Ratio:  4.417              Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 68.750

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-922-340-2910    ..

 Align seg 1/1  to: US-09-922-340-2910  from: 1  to: 424

       5 GlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         ||||||   |||||||||   :::|||   ||||||||||||   |||
     110 CAGTGGTCCTTGCTTCTGGCTGTCCTGGTCTTCTTTCTCTTCGCCTTG 157

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-2909

seq_documentation_block:
; Sequence 2909, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2909
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(477)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-2909


alignment_scores:
            Quality:   53.00            Length:     16
              Ratio:  4.417              Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 68.750

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-922-340-2909    ..

 Align seg 1/1  to: US-09-922-340-2909  from: 1  to: 477

       5 GlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         ||||||   |||||||||   :::|||   ||||||||||||   |||
     110 CAGTGGTCCTTGCTTCTGGCTGTCCTGGTCTTCTTTCTCTTCGCCTTG 157

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-236

 seq_documentation_block:
 ; Sequence 236, Application US/09886055
 ; GENERAL INFORMATION:
 ;  APPLICANT: STRYER, LUBERT
 ;  APPLICANT:  ZOZULYA, SERGEY
 ;  TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
 ;  TITLE OF INVENTION:  BIOSENSORS OF CHEMICAL SENSANTS
 ;  FILE REFERENCE: 078003-0277150
 ;  CURRENT APPLICATION NUMBER: US/09/886,055
 ;  CURRENT FILING DATE:  2001-06-22
 ;  PRIOR APPLICATION NUMBER: 60/213;812
 ;  PRIOR FILING DATE: 2000-06-22
 ;  NUMBER OF SEQ ID NOS: 522
 ;  SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 236
 ;   LENGTH: 945
 ;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-886-055-236


alignment_scores:
            Quality:   53.00            Length:     18
              Ratio:  3.533              Gaps:      0
 Percent Similarity:  83.333   Percent Identity: 44.444

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-886-055-236    ..

 Align seg 1/1  to: US-09-886-055-236  from: 1  to: 945

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         |||:::|||   :::   ||||||||||:::|||:::   ::::::|||::
      76 CCAGAGTGGAAAATGCCCCTGTTCTTGGTGTTCTTGGTGATCTATCTCAT 125

      20 uThr 21
         :|||
     126 CACT 129

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-414

seq_documentation_block:
; Sequence 414, Application US/09886055
; GENERAL INFORMATION:
;  APPLICANT: STRYER, LUBERT
;  APPLICANT:  ZOZULYA, SERGEY
;  TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
;  TITLE OF INVENTION:  BIOSENSORS OF CHEMICAL SENSANTS
;  FILE REFERENCE: 078003-0277150
;  CURRENT APPLICATION NUMBER: US/09/886,055
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: 60/213,812
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 522
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 414
;   LENGTH: 954
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-886-055-414


alignment_scores:
            Quality:   53.00            Length:     18
              Ratio:  3.786              Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 61.111

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-886-055-414    ..

 Align seg 1/1  to: US-09-886-055-414  from: 1  to: 954

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLe 20
         |||          ||||||||||||:::|||||||||   ::::::||||
      61 CCTCCCCTCCAGCTGCTCCTCTTTGTCCTTTTTTTTGCAATTTACCTTCT 110

      20 uThr 21
         ||||
     111 GACA 114

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-969-730-67

 seq_documentation_block:
 ; Sequence 67, Application US/09969730
 ; GENERAL INFORMATION:
 ;  APPLICANT: Ruben et al.
 ;  TITLE OF INVENTION: 90 Human Secreted Proteins
 ;  FILE REFERENCE: PZ013P2
 ;  CURRENT APPLICATION NUMBER: US/09/969,730
 ;  CURRENT FILING DATE:  2001-10-04
```

```
;  PRIOR APPLICATION NUMBER: 09/774,639
;  PRIOR FILING DATE: 2001-02-01
;  PRIOR APPLICATION NUMBER: 60/238,291
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 09/244,112
;  PRIOR FILING DATE: 1999-02-04
;  PRIOR APPLICATION NUMBER: PCT/US98/16235
;  PRIOR FILING DATE: 1998-08-04
;  PRIOR APPLICATION NUMBER: 60/056,371
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,732
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,366
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,364
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,370
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,367
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,365
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,731
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,557
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/056,563
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: 60/055,970
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,986
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,311
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,808
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,803
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,804
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,809
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,806
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,310
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,798
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,309
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,312
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/054,807
;  PRIOR FILING DATE: 1997-08-05
;  PRIOR APPLICATION NUMBER: 60/055,386
;  PRIOR FILING DATE: 1997-08-05
;  NUMBER OF SEQ ID NOS: 373
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
;   LENGTH: 1323
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1086)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1087)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-969-730-67


alignment_scores:
```

```
            Quality:   53.00            Length:     17
              Ratio:  4.077              Gaps:      0
Percent Similarity:  76.471   Percent Identity: 58.824

alignment_block:
 US-09-215-435-179_COPY_1_23 x US-09-969-730-67/rev   ..

 Align seg 1/1  to reverse of: US-09-969-730-67  from: 1  to: 1323

       6 TrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeuThrAr 22
         |||   |||:::|||||||||   ||||||:::||||||   :::   ||
     474 TGGGCCCTTTTCCTCTTCCTCAGTTTCTTCTACCTCTTCAGGTTCCTCAG 425

      22 g 22
         |
     424 G 424
```

OM of: US-09-215-435-179_COPY_1_23 to: EST:* out_format : pfs

Date: Nov 14, 2001 10:30 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_1_23
Query length: 23
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_gss24:AZ337349	-	69.00	154.32	17.77	525	! AZ337349 1M0068D04F Mouse 10kb
gb_est87:BF415978	-	67.00	151.85	24.38	406	! BF415978 UI-R-CA0-bkp-e-12-0-UI
gb_gss31:AZ714249	+	67.00	151.49	25.53	425	! AZ714249 RPCI-24-71J24.TJ RPCI-
gb_gss10:AQ346731	-	67.00	148.84	35.88	595	! AQ346731 RPCI11-125I7.TV RPCI-1
gb_gss20:AZ034197	+	66.00	157.04	12.52	158	! AZ034197 RPCI-23-377D15.TV RPCI
gb_gss20:AZ010385	+	66.00	149.51	32.92	411	! AZ010385 RPCI-23-365G17.TJ RPCI
gb_gss28:AZ556218	+	66.00	149.30	33.81	422	! AZ556218 RPCI-23-177P7.TV RPCI-
gb_est30:AU039151	-	65.00	153.30	20.23	191	! AU039151 AU039151 Dictyostelium
gb_est21:AI490726	-	65.00	149.14	34.52	324	! AI490726 EST241434 tomato shoot
gb_est21:AI490728	+	65.00	148.39	37.97	356	! AI490728 EST241436 tomato shoot
gb_est97:BG129260	-	65.00	147.24	44.01	412	! BG129260 EST474906 tomato shoot
gb_est97:BG130556	-	65.00	147.22	44.12	413	! BG130556 EST476202 tomato shoot
gb_est48:AW560372	+	65.00	145.26	56.78	530	! AW560372 EST315420 DSIR Medicag
gb_gss13:AQ554809	-	65.00	145.20	57.21	534	! AQ554809 RPCI-11-382G13.TJ RPCI
gb_est103:C91466	-	65.00	144.71	60.89	568	! C91466 C91466 Dictyostelium dis
gb_est80:BE910551	-	65.00	143.76	68.81	641	! BE910551 601501101F1 NIH_MGC_70
gb_est93:BF867577	-	64.50	139.07	125.59	1008	! BF867577 963092F09.y1 C. reinh
gb_est71:BE158828	+	64.00	144.83	60.00	421	! BE158828 IL2-HT0397-151299-026-
gb_gss34:AZ904272	-	64.00	144.48	62.74	440	! AZ904272 RPCI-24-221F13.TJ RPCI
gb_est22:AI616040	+	64.00	144.04	66.34	465	! AI616040 mq68b08.y1 Soares_thym
gb_gss18:AQ936195	+	64.00	143.09	75.00	525	! AQ936195 RPCI-11-Seg1-4-608P20-
gb_gss25:AZ373346	+	64.00	142.44	81.51	570	! AZ373346 1M0125O05R Mouse 10kb
gb_gss26:AZ409528	+	64.00	142.43	81.65	571	! AZ409528 1M0181L11F Mouse 10kb
gb_gss32:AZ790560	+	64.00	140.62	102.93	718	! AZ790560 2M0039E12F Mouse 10kb
gb_est76:BE615675	-	64.00	140.23	108.15	754	! BE615675 601279360F1 NIH_MGC_39
gb_est9:AA560483	+	63.00	146.08	51.08	270	! AA560483 v131f10.r1 Stratagene
gb_est108:T07865	-	63.00	146.08	51.08	270	! T07865 EST05755 Fetal brain, St
gb_est20:AI437332	+	63.00	143.35	72.54	382	! AI437332 fb30b04.x1 Zebrafish W
gb_gss29:AZ619755	+	63.00	140.92	99.08	520	! AZ619755 1M0452I08F Mouse 10kb
gb_gss27:AZ477895	+	63.00	139.37	120.87	633	! AZ477895 1M0297C04R Mouse 10kb
gb_gss1:AA550316	-	63.00	137.99	144.25	754	! AA550316 1457m3 gmbPfHB3.1, G.
gb_est103:C23639	-	63.00	136.31	178.91	933	! C23639 C23639 Dictyostelium dis
gb_gss16:AQ750720	-	63.00	135.90	188.60	983	! AQ750720 HS_5574_B2_G03_T7A RPC
gb_gss16:AQ780342	-	63.00	134.36	229.57	1194	! AQ780342 HS_3184_A2_D07_T7C CI
em_estpl9:BF262994	-	63.00	133.44	258.36	1342	! Bf262994 HV_CEa0005I08f Hordeu
gb_est37:AV356560	-	62.50	149.39	33.44	154	! AV356560 AV356560 RIKEN full-le
gb_est28:AL378550	-	62.50	140.44	105.31	479	! AL378550 MtBB38H07F1 MtBB Medic
gb_est90:BF645988	-	62.50	139.03	126.22	573	! BF645988 NF065F02EC1F1026 Elici
gb_est89:BF520412	-	62.50	138.93	127.78	580	! BF520412 EST457882 DSIL Medicag
gb_gss27:AZ474525	+	62.50	138.65	132.46	601	! AZ474525 1M0292M02F Mouse 10kb
gb_est73:BE325277	-	62.50	138.01	143.82	652	! BE325277 NF119D06ST1F1057 Devel
gb_gss29:AZ611033	-	62.50	137.26	158.33	717	! AZ611033 1M0436L02R Mouse 10
gb_est15:AI066195	+	62.00	147.82	40.90	163	! AI066195 TENU2586 T. cruzi e
gb_gss21:AZ100984	-	62.00	144.80	60.22	239	! AZ100984 RPCI-23-480N14.TV R
gb_est100:BG426365	-	62.00	138.86	129.06	508	! BG426365 602493552F1 NIH_MGC

seq_name: gb_gss24:AZ337349

seq_documentation_block:

```
LOCUS       AZ337349      525 bp    DNA              GSS       29-SEP-2000
DEFINITION  1M0068D04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0068D04 F, DNA sequence.
ACCESSION   AZ337349
VERSION     AZ337349.1  GI:10407556
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0068  row: D  column: 04
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0068D04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      126 a    146 c    157 g     95 t      1 others
ORIGIN
```

alignment_scores:
Quality: 69.00 Length: 22

```
            Ratio:  3.450              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-215-435-179_COPY_1_23 x AZ337349/rev   ..

 Align seg 1/1  to reverse of: AZ337349  from: 1  to: 525

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         :::   |||||||:::::::::|||||||||||||||||||||:::|||||
     384 CTCTATCTCCCCCGCTATTTCCTCCTCCTCTTCCTCCTCTTCCTCTTTCT 335

      17 uPheLeuLeuThrArg 22
         |:::|||||||   |||
     334 CCTCCTCCTCCTTCGC 319

seq_name: gb_est87:BF415978

seq_documentation_block:
LOCUS       BF415978      406 bp    mRNA            EST       28-NOV-2000
DEFINITION  UI-R-CA0-bkp-e-12-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
            UI-R-CA0-bkp-e-12-0-UI 3', mRNA sequence.
ACCESSION   BF415978
VERSION     BF415978.1  GI:11403967
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized corpus-striatum library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 3-129, >RMER1B#Unknown
            130-166, >(GGA)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA0-bkp-e-12-0-UI"
                     /clone_lib="UI-R-CA0"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain,  cerebral cortex, corpus striatum, testis, and
                     hippocampus.   For a detailed description of the library
                     from which this clone was  derived, please visit our web
                     site at ratest.eng.uiowa.edu.   The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_LIB=UI-R-CA0
                     TAG_TISSUE=corpus-striatum
                     TAG_SEQ=CTAGG"
BASE COUNT      116 a    102 c     74 g    114 t
ORIGIN


alignment_scores:
            Quality:   67.00              Length:     27
              Ratio:  3.045                Gaps:      1
 Percent Similarity:  81.481   Percent Identity: 55.556

alignment_block:
 US-09-215-435-179_COPY_1_23 x BF415978/rev   ..

 Align seg 1/1  to reverse of: BF415978  from: 1  to: 406

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ::::::|||||   :::|||||||:::|||:::|||||||||||||||||
     161 CTCCTCCTCCCTCTCTTCCTCCTGATACTCCTTCTCCTCTTCTTCTTCCT 112

      17 uPheLeu...........LeuThrArgGly 23
         |:::|||            :::||||||||||
     111 CCTTCTTCATTTTTCGGAGGTCACAAGGGGA 81

seq_name: gb_gss31:AZ714249

seq_documentation_block:
LOCUS       AZ714249      425 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-71J24.TJ RPCI-24 Mus musculus genomic clone RPCI-24-71J24,
            DNA sequence.
ACCESSION   AZ714249
VERSION     AZ714249.1  GI:12449772
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 71  row: J  column: 24
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .425
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-71J24"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
```

```
BASE COUNT      119 a     96 c    103 g    107 t
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     16
              Ratio:  4.188               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AZ714249   ..

 Align seg 1/1  to: AZ714249  from: 1  to: 425

       5 GlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
         :::|||||||||||||||||||||:::|||||||||:::|||:::|||
     376 GAGTGGCTCCTCCTTCTTTTTCTTTTCTTTTTTTTTTTTTTTTTCTT 423

seq_name: gb_gss10:AQ346731

seq_documentation_block:
LOCUS       AQ346731      595 bp    DNA             GSS       07-MAY-1999
DEFINITION  RPCI11-125I7.TV RPCI-11 Homo sapiens genomic clone RPCI-11-125I7,
            DNA sequence.
ACCESSION   AQ346731
VERSION     AQ346731.1  GI:4171627
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-125I7.TJ
            Contact: Shaying Zhao; William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Homo sapiens"
                     /db_xref="GDB:7547814"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-125I7"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      199 a    124 c    130 g    141 t      1 others
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     16
              Ratio:  4.467               Gaps:      0
 Percent Similarity:  93.750   Percent Identity: 75.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AQ346731/rev   ..

 Align seg 1/1  to reverse of: AQ346731  from: 1  to: 595

       4 ProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeu 19
         |||   ||||||||||||||||||:::::|||||||||:::|||||||
     593 CCAACTTGGCTCCTCCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTG 546

seq_name: gb_gss20:AZ034197

seq_documentation_block:
LOCUS       AZ034197      158 bp    DNA             GSS       01-MAR-2000
DEFINITION  RPCI-23-377D15.TV RPCI-23 Mus musculus genomic clone RPCI-23-377D15
            , DNA sequence.
ACCESSION   AZ034197
VERSION     AZ034197.1  GI:7119391
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 158)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-377D15.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 377  row: D  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .158
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-377D15"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT        2 a     70 c      3 g     83 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     14
              Ratio:  4.714               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AZ034197   ..
```

```
Align seg 1/1  to: AZ034197  from: 1  to: 158

      7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
        ||||||||||||||||||||||||||||||||||||||||||||
     67 CTTCTTCTCCTCTTCCTCCTTTTTTTCTTCCTCTTCCTCCTC 108

seq_name: gb_gss20:AZ010385

seq_documentation_block:
LOCUS       AZ010385      411 bp    DNA             GSS       25-FEB-2000
DEFINITION  RPCI-23-365G17.TJ RPCI-23 Mus musculus genomic clone RPCI-23-365G17
            , DNA sequence.
ACCESSION   AZ010385
VERSION     AZ010385.1  GI:7085769
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 365  row: G  column: 17
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-365G17"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       49 a    153 c     40 g    169 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     14
              Ratio:  4.714               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AZ010385   ..

 Align seg 1/1  to: AZ010385  from: 1  to: 411

      7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
        ||||||||||||||||||||||||||||||||||||||||||||
```

```
     75 CTCCTTCTCCTCTTCCTCCTTTTCTTCTTCCTCTTCCTCCTC 116

seq_name: gb_gss28:AZ556218

seq_documentation_block:
LOCUS       AZ556218      422 bp    DNA             GSS       20-NOV-2000
DEFINITION  RPCI-23-177P7.TV RPCI-23 Mus musculus genomic clone RPCI-23-177P7,
            DNA sequence.
ACCESSION   AZ556218
VERSION     AZ556218.1  GI:11236038
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-177P7.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 177  row: P  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-177P7"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       87 a    119 c     92 g    124 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     14
              Ratio:  4.714               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AZ556218   ..

 Align seg 1/1  to: AZ556218  from: 1  to: 422

      7 LeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLeuLeu 20
        ||||||||||||||||||||||||||||||||||||||||||||
     23 CTTCTTCTCCTCTTCCTCCTTTTTTTCTTCCTCTTCCTCCTC 64

seq_name: gb_est30:AU039151
```

```
seq_documentation_block:
LOCUS       AU039151      191 bp    mRNA            EST       29-MAR-1999
DEFINITION  AU039151 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSM450, mRNA sequence.
ACCESSION   AU039151
VERSION     AU039151.1  GI:3985904
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 191)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .191
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSM450"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      130 a      8 c      8 g     45 t
ORIGIN


alignment_scores:
           Quality:   65.00             Length:     18
             Ratio:  4.062               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 72.222

alignment_block:
 US-09-215-435-179_COPY_1_23 x AU039151/rev   ..

 Align seg 1/1  to reverse of: AU039151  from: 1  to: 191

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         :::|||      |||||||||||||||:::|||||||||||||||:::||
      97 TTACTATATTACTGGTTATTATTATTATTATTATTTTTTTTTTTTTT 48

      18 eLeu 19
         ||||
      47 TTTA 44

seq_name: gb_est21:AI490726

seq_documentation_block:
LOCUS       AI490726      324 bp    mRNA            EST       29-JUN-1999
DEFINITION  EST241434 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
            cLEB3L10, mRNA sequence.
ACCESSION   AI490726
VERSION     AI490726.1  GI:4386036
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   van der Hoeven,R.S., Matern,A.L., Vision,T., Holt,I.E., Liang,F.,
            Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
            and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato shoot meristem
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: EST241436
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEB3L10"
                     /clone_lib="tomato shoot, Cornell"
                     /tissue_type="shoot meristem"
                     /dev_stage="8 week old plants"
                     /lab_host="XLOLR"
                     /note="Vector: pBK_CMV; Site_1: EcoR1; Site_2: Xho1; cLEB
                     - Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
                     library  made from tomato vegetative shoots including
                     meristems and small  expanding leaves."
BASE COUNT      103 a     43 c    105 g     73 t
ORIGIN


alignment_scores:
           Quality:   65.00             Length:     19
             Ratio:  3.824               Gaps:      0
 Percent Similarity:  89.474   Percent Identity: 57.895

alignment_block:
 US-09-215-435-179_COPY_1_23 x AI490726/rev   ..

 Align seg 1/1  to reverse of: AI490726  from: 1  to: 324

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         ::::::   :::||||||:::|||:::|||||||||   ||||||||||
     176 CTTTTAGCAGGTGGTTGATCCTCGTCTTCCTCCTCGTCTTCTTCCTCTT 127

      18 eLeuLeu 20
         ||||:::
     126 CCTCATC 120

seq_name: gb_est21:AI490728

seq_documentation_block:
LOCUS       AI490728      356 bp    mRNA            EST       29-JUN-1999
DEFINITION  EST241436 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
            cLEB3L10, mRNA sequence.
ACCESSION   AI490728
VERSION     AI490728.1  GI:4386038
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   van der Hoeven,R.S., Matern,A.L., Vision,T., Holt,I.E., Liang,F.,
            Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
            and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato shoot meristem
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: EST241434
            Contact: David Frisch
```

```
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 4366
          Fax: 864 656 4293
          Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEB3L10"
                     /clone_lib="tomato shoot, Cornell"
                     /tissue_type="shoot meristem"
                     /dev_stage="8 week old plants"
                     /lab_host="XLOLR"
                     /note="Vector: pBK_CMV; Site_1: EcoR1; Site_2: Xho1; cLEB
                     - Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
                     library  made from tomato vegetative shoots including
                     meristems and small  expanding leaves."
BASE COUNT       91 a    111 c     46 g    108 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     19
              Ratio:  3.824               Gaps:       0
 Percent Similarity:  89.474   Percent Identity: 57.895

alignment_block:
 US-09-215-435-179_COPY_1_23 x AI490728   ..

 Align seg 1/1  to: AI490728  from: 1  to: 356

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         ::::::   :::|||||:::|||:::|||||||||   ||||||||||
     181 CTTTTTAGCAGGTGGTTGATCCTCGTCTTCCTCCTCGTCTTCTTCCTCTT 230

      18 eLeuLeu 20
         ||||:::
     231 CCTCATC 237

seq_name: gb_est97:BG129260

seq_documentation_block:
LOCUS        BG129260      412 bp    mRNA            EST       31-JAN-2001
DEFINITION   EST474906 tomato shoot/meristem Lycopersicon esculentum cDNA clone
             cTOF23F6 5' sequence, mRNA sequence.
ACCESSION    BG129260
VERSION      BG129260.1  GI:12629448
KEYWORDS     EST.
SOURCE       tomato.
  ORGANISM   Lycopersicon esculentum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
             Lycopersicon.
REFERENCE    1  (bases 1 to 412)
  AUTHORS    van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
             Hansen,C., Ronning,C. and Tanksley,S.
  TITLE      Generation of ESTs from tomato shoot/meristem tissue
  JOURNAL    Unpublished (2001)
COMMENT      Contact: CUGI
             Clemson University Genomics Institute
             Clemson University
             100 Jordan Hall, Clemson, SC 29634, USA
             Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOF23F6"
                     /clone_lib="tomato shoot/meristem"
                     /tissue_type="shoot/meristem"
                     /dev_stage="developing shoots from 4-6wks old plants"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Small expanding leaves from the growing tip were
                     taken from greenhouse plants (4-6wks old TA496). Tissue
                     was immediately frozen in liquid nitrogen."
BASE COUNT       128 a     53 c    118 g    113 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     19
              Ratio:  3.824               Gaps:       0
 Percent Similarity:  89.474   Percent Identity: 57.895

alignment_block:
 US-09-215-435-179_COPY_1_23 x BG129260/rev   ..

 Align seg 1/1  to reverse of: BG129260  from: 1  to: 412

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         ::::::   :::|||||:::|||:::|||||||||   ||||||||||
     182 CTTTTTAGCAGGTGGTTGATCCTCGTCTTCCTCCTCGTCTTCTTCCTCTT 133

      18 eLeuLeu 20
         ||||:::
     132 CCTCATC 126

seq_name: gb_est97:BG130556

seq_documentation_block:
LOCUS        BG130556      413 bp    mRNA            EST       31-JAN-2001
DEFINITION   EST476202 tomato shoot/meristem Lycopersicon esculentum cDNA clone
             cTOF30P11 5' sequence, mRNA sequence.
ACCESSION    BG130556
VERSION      BG130556.1  GI:12630744
KEYWORDS     EST.
SOURCE       tomato.
  ORGANISM   Lycopersicon esculentum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
             Lycopersicon.
REFERENCE    1  (bases 1 to 413)
  AUTHORS    van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
             Hansen,C., Ronning,C. and Tanksley,S.
  TITLE      Generation of ESTs from tomato shoot/meristem tissue
  JOURNAL    Unpublished (2001)
COMMENT      Contact: CUGI
             Clemson University Genomics Institute
             Clemson University
             100 Jordan Hall, Clemson, SC 29634, USA
             Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOF30P11"
                     /clone_lib="tomato shoot/meristem"
                     /tissue_type="shoot/meristem"
                     /dev_stage="developing shoots from 4-6wks old plants"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Small expanding leaves from the growing tip were
                     taken from greenhouse plants (4-6wks old TA496). Tissue
                     was immediately frozen in liquid nitrogen."
BASE COUNT       128 a     53 c    118 g    114 t
ORIGIN
```

```
alignment_scores:
            Quality:   65.00             Length:     19
              Ratio:  3.824               Gaps:      0
 Percent Similarity:  89.474   Percent Identity: 57.895

alignment_block:
 US-09-215-435-179_COPY_1_23 x BG130556/rev   ..

 Align seg 1/1  to reverse of: BG130556  from: 1  to: 413

       2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
         ::::::   :::|||||||:::|||:::|||||||||   ||||||||||
     182 CTTTTTAGCAGGTGGTTGATCCTCGTCTTCCTCCTCGTCTTCTTCCTCTT 133

      18 eLeuLeu 20
         ||||:::
     132 CCTCATC 126

seq_name: gb_est48:AW560372

seq_documentation_block:
LOCUS       AW560372      530 bp    mRNA            EST       07-SEP-2000
DEFINITION  EST315420 DSIR Medicago truncatula cDNA clone pDSIR-26N4, mRNA
            sequence.
ACCESSION   AW560372
VERSION     AW560372.1  GI:7205798
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after inoculation with
            Phytophthora medicaginis
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Minnesota EST name:M251727e ; TIGR sequence name:MTBAX74TK ; More
            information, including clone ordering, is available at. .
            'http://chrysie.tamu.edu/medicago'
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pDSIR-26N4"
                     /clone_lib="DSIR"
                     /tissue_type="roots infected with Phytophthora
                     medicaginis"
                     /dev_stage="roots harvested at 10 days post inoculation
                     with Phytophthora medicaginis"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10 days post inoculation with
                     Phytophthora medicaginis. The cDNA was directionally
                     ligated into the Uni-ZAP XR vector from Stratagene and
                     packaged using Gigapack III Gold packaging extracts.
                     Plasmids containing cDNA inserts were excised from the
                     recombinant lambda-ZAP phage using Ex-Assist helper phage
                     and propagated in XLOLR cells. Note: EST may be of fungal
                     origin."
BASE COUNT       86 a    147 c     39 g    258 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     20
              Ratio:  3.421               Gaps:      0
 Percent Similarity:  95.000   Percent Identity: 65.000

alignment_block:
 US-09-215-435-179_COPY_1_23 x AW560372   ..

 Align seg 1/1  to: AW560372  from: 1  to: 530

       1 MetMetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLe 17
         ::::::|||   ::::::|||||||||||||||||||:::||||||||||
     236 CTTATCCTTTCCAGATTTCTTCTTCTTCTCTTCCTTGTCTTCTTTTTTCT 285

      17 uPheLeuLeu 20
         |:::||||||
     286 TCTTCTTCTC 295

seq_name: gb_gss13:AQ554809

seq_documentation_block:
LOCUS       AQ554809      534 bp    DNA             GSS       28-MAY-1999
DEFINITION  RPCI-11-382G13.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-382G13
            , DNA sequence.
ACCESSION   AQ554809
VERSION     AQ554809.1  GI:4913986
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-382G13.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Homo sapiens"
                     /db_xref="GDB:7646460"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-382G13"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      216 a     64 c    141 g    113 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     19
```

```
            Ratio:  3.824                Gaps:      0
 Percent Similarity:  89.474   Percent Identity: 68.421

alignment_block:
 US-09-215-435-179_COPY_1_23 x AQ554809/rev   ..

 Align seg 1/1  to reverse of: AQ554809  from: 1  to: 534

       3 LeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPheLe 19
         ||||||:::||||||||||||   :::|||   :::||||||||||||::
     226 TTACCTCACTGGTTACTCCTGGATCTCCTTTGCTGGTTCTTCCTCTTCTT 177

      19 uLeuThr 21
         :||||||
     176 CCTGACA 170

seq_name: gb_est103:C91466

seq_documentation_block:
LOCUS       C91466         568 bp    mRNA            EST       20-APR-1998
DEFINITION  C91466 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSK314, mRNA sequence.
ACCESSION   C91466
VERSION     C91466.1  GI:3060832
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Yoshino,R., Morio,T. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSK314"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      242 a      94 c      67 g     160 t       5 others
ORIGIN


 alignment_scores:
             Quality:   65.00              Length:     18
               Ratio:  4.062                Gaps:      0
  Percent Similarity:  88.889   Percent Identity: 72.222

 alignment_block:
  US-09-215-435-179_COPY_1_23 x C91466/rev   ..

  Align seg 1/1  to reverse of: C91466  from: 1  to: 568

        2 MetLeuProGlnTrpLeuLeuLeuLeuPheLeuLeuPhePhePheLeuPh 18
          :::|||      ||||||||||||||||:::||||||||||||||||:::||
      485 TTACTATATTACTGGTTATTATTATTATTATTATTTTTTTTTTTTTTT 436

       18 eLeu 19
          ||||
      435 TTTA 432
```

```
OM of: US-09-215-435-179_COPY_24_121 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:47 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pr6:AL157823         +  343.00  618.55  3.2e-26   127384  ! AL157823 Human DNA sequence
gb_htg22:AL512371       +  343.00  616.69  4.1e-26   159482  ! AL512371 Homo sapiens chromc
gb_htg23:AL590400       +  343.00  616.21  4.4e-26   168884  ! AL590400 Homo sapiens chromc
gb_om:CCCOLIP           +  147.00  281.66  1.9e-07   452  ! X53564 Canine mRNA for colipase
gb_om:RABCOLIPAS        +  141.00  269.79  8.6e-07   459  ! L06329 Oryctolagus cuniculus pa
gb_pr10:HUMCOLIP        +  135.00  256.98  4.5e-06   523  ! J02883 Human colipase mRNA, com
gb_pat1:AX014220        +  135.00  255.10  5.7e-06   656  ! AX014220 Sequence 16 from Paten
gb_ro1:MCCOLIPAS        +  134.50  256.54  4.7e-06   490  ! X82998 M.coypus mRNA for colipa
gb_om:ECCOLIPB          +  134.00  255.44  5.4e-06   497  ! X74344 E.caballus mRNA for coli
gb_om:AF148567          +  133.00  253.27  7.2e-06   510  ! AF148567 Sus scrofa pancreatic
gb_om:ECCO1IPA          +  131.00  250.00  1.1e-05   472  ! X74503 E.caballus mRNA for coli
gb_ro2:RATCOLIP         +  122.00  232.85  9.8e-05   446  ! M33333 Rat colipase mRNA, compl
gb_ro2:RATCOLQ          +  122.00  232.04  0.0001    492  ! M58370 Rat colipase mRNA, compl
gb_in3:CEW01B6          +   90.00  135.98  24.49     28122  ! Z68301 Caenorhabditis elegans
gb_pr10:HSY10344        -   87.00  149.38  4.39      2734  ! Y10344 H.sapiens GalNAc-T2 gen
gb_htg19:AL136988       +   87.00  116.46  299.67    146755  ! AL136988 Homo sapiens chromc
gb_htg9:AC024008        +   87.00  115.52  337.90    164352  ! AC024008 Homo sapiens chromc
gb_htg13:AC067715       +   87.00  114.23  398.51    192027  ! AC067715 Mus musculus chromc
gb_ba3:MTCY3C7          +   86.50  127.26  74.98     35292  ! Z82098 Mycobacterium tubercul
gb_in3:CEY47H9C         -   84.50  117.08  276.78    75342  ! AL032657 Caenorhabditis elega
gb_htg24:CEY47H9_4      -   84.50  116.30  305.70    82746  ! Continuation (5 of 5) of CEY4
gb_pr5:AK025845         -   83.00  145.32  7.39      1733  ! AK025845 Homo sapiens cDNA: FI
gb_pat1:A96992          +   82.50  151.07  3.54      768  ! A96992 Sequence 4 from Patent W
gb_pr1:AB035180S2       +   82.50  151.04  3.55      771  ! AB035181 Homo sapiens Dickkopf-
gb_pat1:A96990          +   82.50  149.94  4.09      881  ! A96990 Sequence 2 from Patent W
gb_pr9:HSA406946        +   82.50  148.55  4.89      1042  ! AJ406946 Homo sapiens mRNA for
gb_pr1:AB033208         +   82.50  147.84  5.35      1135  ! AB033208 Homo sapiens dickkopf
gb_pr1:AB033941         +   82.50  145.68  7.06      1475  ! AB033941 Homo sapiens mRNA for
gb_ro1:MMU243963        +   82.50  141.39  12.24     2478  ! AJ243963 Mus musculus mRNA for
gb_pr5:AF177395         +   82.50  138.14  18.57     3671  ! AF177395 Homo sapiens dickkopf
gb_pl1:AB005237         -   82.50  111.89  537.96    87835  ! AB005237 Arabidopsis thaliana
gb_htg23:AP001863       -   82.50  111.39  574.15    93398  ! AP001863 Homo sapiens chromos
gb_pr1:AC003099         -   82.50  111.23  585.44    95129  ! AC003099 Homo sapiens chromos
gb_pr7:AP001821         -   82.50  110.16  672.30    108388  ! AP001821 Homo sapiens genomi
gb_pr2:AC006070         -   82.50  106.83  1.0e+03   161987  ! AC006070 Homo sapiens chromc
gb_htg10:AC025904       +   82.50  106.24  1.1e+03   174032  ! AC025904 Homo sapiens clone
gb_htg14:AC068842       -   82.50  105.79  1.2e+03   183719  ! AC068842 Homo sapiens chromc
gb_pr9:HSA406951        +   82.00  148.31  5.04      953  ! AJ406951 Homo sapiens mRNA for
gb_htg18:AC085286       -   82.00  148.27  5.07      958  ! AC085286 Giardia intestinalis c
gb_ro2:RRU25680         +   82.00  146.74  6.16      1152  ! U25680 Rattus norvegicus pread
gb_ro1:D84336           +   82.00  144.35  8.37      1538  ! D84336 Rattus norvegicus ZOG m
gb_in2:AF226726         -   82.00  143.58  9.25      1689  ! AF226726 Giardia intestinal
gb_om:DOGCOLIP          +   82.00  138.39  17.99     3164  ! M63427 Dog pancreatic colip
gb_htg24:AP003455       -   82.00  104.39  1.4e+03   193530  ! AP003455 Oryza sativa chr
gb_pr4:AC026888         -   82.00  103.75  1.5e+03   208894  ! AC026888 Homo sapiens chr

seq_name: gb_pr6:AL157823

seq_documentation_block:
LOCUS       AL157823   127384 bp    DNA             PRI       10-FEB-2001
DEFINITION  Human DNA sequence from clone RP3-510O8 on chromosome 6, complete
            sequence.
ACCESSION   AL157823
VERSION     AL157823.9  GI:12750800
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127384)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 12, 2001 this sequence version replaced gi:12709869.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-510O8 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-510O8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP3-510O8 is at 1 in this sequence. The
            true left end of clone RP3-422H11 is at 127285 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .127384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-510O8"
                     /clone_lib="RPCI-3"
     misc_feature    6404
                     /note="Weak data in alu."
     misc_feature    complement(95875. .96037)
                     /note="match: STS: Em:G13580"
     misc_feature    95894. .96129
                     /note="match: STS: Em:G14863"
BASE COUNT    33605 a  30130 c  31408 g  32241 t
ORIGIN
```

```
alignment_scores:
            Quality:  343.00              Length:    340
              Ratio:  3.811                Gaps:      2
 Percent Similarity:  26.471   Percent Identity: 25.882

alignment_block:
 US-09-215-435-179_COPY_24_121 x AL157823   ..

 Align seg 1/1  to: AL157823  from: 1  to: 127384

       8 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 24
         :::|||   ||||||||||||||||||||||||||||||||||||||||
   87881 TCCCTGCAGGAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGAC 87930

      24 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   87931 TGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGA 87980

      41 ysGlySerGluGlySerLeuCysGlnThrGlnVal..Phe..... 53
         ||||||||||||||||||||||||||||||||||  :::
   87981 AGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGGGTATCGCCGCCCGG 88030

      53 ..................................................... 53

   88031 GGGGAGCCAGAGGGGATCCAGGGGAAGTGGGAGCCAGGGCGGGCCCAGAT 88080

      53 ..................................................... 53

   88081 TCCTGGGGAGGAAAGAAGGGTAGGTGGGCGGAAATGCCCTGGAGCCTGGA 88130

      53 ..................................................... 53

   88131 ATTCCCCCGTGGGACTGCACCTCCAAGCCTTCTGGGGGAGCCTGTGGACG 88180

      53 ..................................................... 53

   88181 GAACCAATATGGTAGCCCACCCTAGGGTTTTGGAATCCTGGATTGGGGCG 88230

      53 ..................................................... 53

   88231 GGAAGGCGATGTTTCACTTTCTAATGTGGGTTCACATTCTCCAGGAGTCT 88280

      53 ..................................................... 53

   88281 TCTCAGCAAATTCCGAAAGTAGTACAGTAAATGGGCACAGCTAGGTCCAG 88330

      53 ..................................................... 53

   88331 CCTCTGCCCCAGAGCCCCGTCTTCATCCAGCTCAGGAGTTCTGCTGTGGC 88380

      53 ..................................................... 53

   88381 CCCGGGTGTTCCTGGTGGAGGATGGGCAGAGGCAGCCTCTGAGCTGGGTG 88430

      53 ..................................................... 53

   88431 TGTAGGGGAGGGAGTTGATTTGCCCATGGAAGGATTTAAAGACAAGACAT 88480

      53 ..................................................... 53

   88481 AGAGGTCCTGCGATGGTGTGTCCATGGACTTAGGACAGGCAGATGGGTCA 88530

      53 ..................................................... 53

   88531 GTGACAGGCAGGGTTACAAGGTCTGGAGCTGGGAGGACACTGAACAATCC 88580

      53 ..................................................... 53

   88581 AGGAGCCAGCCTGGAGAAGGCAGCCAGAGGTGCCACTCTGAGCAGCTGGC 88630

      53 ..................................................... 53

   88631 AAGCTATCATTTATCCCTGGATCCTGGCTGTGGGCAAGCACATGGGCCCT 88680

      53 ..................................................... 53

   88681 GGAGCTGGGCTGGGCTGGGCTTAGGGGCTCAGGGAAGAGCTGCCGCCAAG 88730

      54 ..............................PheGlyGlnTyrArgA 59
                                      ||||||||||||||||
   88731 GTCGAGGTCAAAGTCCTGTCTTTCCCAGGTGTTCTTTGGCCAATATAGAG 88780

      59 laCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLys 75
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   88781 CGTGTCCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAA 88830

      76 TrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLe 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   88831 TGGCTTAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTT 88880

      92 uAlaLysLysMetPhePhe 98
         ||||||||||||||||||
   88881 GGCTAAGAAAATGTTCTTC 88899

seq_name: gb_htg22:AL512371

seq_documentation_block:
LOCUS       AL512371   159482 bp    DNA             HTG       21-MAR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-238F24, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL512371
VERSION     AL512371.4  GI:13274864
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159482)
  AUTHORS   Clark,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 12, 2001 this sequence version replaced gi:12964529.
            ------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ------------- Project Information
            Center project name: bA238F24
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 154970 bases at least Q40
            Consensus quality: 156246 bases at least Q30
            Consensus quality: 156738 bases at least Q20
            Insert size: 159282; sum-of-contigs
            Insert size: 165487; 7.5% error; agarose-fp
            Quality coverage: 6.35x in Q20 bases; sum-of-contigs Quality
            coverage: 6.11x in Q20 bases; agarose-fp
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    3066: contig of 3066 bp in length
            *    3067 3166: gap of      100 bp
            *    3167  155002: contig of 151836 bp in length
            *  155003 155102: gap of      100 bp
            *  155103  159482: contig of 4380 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .159482
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-238F24"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .3066
                     /note="assembly_fragment:02384
                     fragment_chain:1"
     misc_feature    3167. .155002
                     /note="assembly_fragment:02553
                     fragment_chain:1"
     misc_feature    155103. .159482
                     /note="assembly_fragment:02886"
BASE COUNT    42530 a  37388 c  38532 g  40831 t    201 others
ORIGIN


alignment_scores:
            Quality:  343.00             Length:    340
              Ratio:  3.811                Gaps:      2
 Percent Similarity:  26.471   Percent Identity: 25.882

alignment_block:
 US-09-215-435-179_COPY_24_121 x AL512371   ..

 Align seg 1/1  to: AL512371  from: 1  to: 159482

       8 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 24
         :::|||   ||||||||||||||||||||||||||||||||||||||||
  101423 TCCCTGCAGGAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGAC 101472

      24 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 41
         |||||||||||||||||||||||||||||||||||||||||||||||||
  101473 TGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGA 101522

      41 ysGlySerGluGlySerLeuCysGlnThrGlnVal..Phe..... 53
         |||||||||||||||||||||||||||||||||  :::
  101523 AGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGGGTATCGCCGCCCGG 101572

      53 ..................................................... 53

  101573 GGGGAGCCAGAGGGGATCCAGGGGAAGTGGGAGCCAGGGAGGGCCCAGAT 101622

      53 ..................................................... 53

  101623 TCCTGGGGAGGAAAGAAGGGTAGGTGGGCGGAAATGCCCTGGAGCCTGGA 101672

      53 ..................................................... 53

  101673 ATTCCCCCGTGGGACTGCACCTCCAAGCCTTCTGGGGGAGCCTGTGGACG 101722

      53 ..................................................... 53

  101723 GAACCAATATGGTAGCCCACCCTAGGGTTTTGGAATCCTGGATTGGGGCG 101772

      53 ..................................................... 53

  101773 GGAAGGCGATGTTTCACTTTCTAATGTGGGTTCACATTCTCCAGGAGTCT 101822

      53 ..................................................... 53

  101823 TCTCAGCAAATTCCGAAAGTAGTACAGTAAATGGGCACAGCTAGGTCCAG 101872

      53 ..................................................... 53

  101873 CCTCTGCCCCAGAGCCCCGTCTTCATCCAGCTCAGGAGTTCTGCTGTGGC 101922

      53 ..................................................... 53

  101923 CCCGGGTGTTCCTGGTGGAGGATGGGCAGAGGCAGCCTCTGAGCTGGGTG 101972

      53 ..................................................... 53

  101973 TGTAGGGGAGGGAGTTGATTTGCCCATGGAAGGATTTAAAGACAAGACAT 102022

      53 ..................................................... 53

  102023 AGAGGTCCTGCGATGGTGTGTCCATGGACTTAGGACAGGCAGATGGGTCA 102072

      53 ..................................................... 53

  102073 GTGACAGGCAGGGTTACAAGGTCTGGAGCTGGGAGGACACTGAACAATCC 102122

      53 ..................................................... 53

  102123 AGGAGCCAGCCTGGAGAAGGCAGCCAGAGGTGCCACTCTGAGCAGCTGGC 102172

      53 ..................................................... 53

  102173 AAGCTATCATTTATCCCTGGATCCTGGCTGTGGGCAAGCACATGGGCCCT 102222

      53 ..................................................... 53

  102223 GGAGCTGGGCTGGGCTGGGCTTAGGGGCTCAGGGAAGAGCTGCCGCCAAG 102272

      54 ..............................PheGlyGlnTyrArgA 59
                                      ||||||||||||||||
  102273 GTCGAGGTCAAAGTCCTGTCTTTCCCAGGTGTTCTTTGGCCAATATAGAG 102322

      59 laCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLys 75
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  102323 CGTGTCCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAA 102372

      76 TrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLe 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  102373 TGGCTTAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTT 102422

      92 uAlaLysLysMetPhePhe 98
         |||||||||||||||||||
  102423 GGCTAAGAAAATGTTCTTC 102441

seq_name: gb_htg23:AL590400

seq_documentation_block:
LOCUS       AL590400   168884 bp    DNA             HTG       08-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-569I12, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   AL590400
VERSION     AL590400.2  GI:13568356
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168884)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:13559131.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA569I12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 164511 bases at least Q40
            Consensus quality: 165952 bases at least Q30
```

```
        Consensus quality: 166791 bases at least Q20
        Insert size: 167584; sum-of-contigs
        Insert size: 185239; 2.7% error; agarose-fp
        Quality coverage: 6.19x in Q20 bases; sum-of-contigs Quality
        coverage: 5.74x in Q20 bases; agarose-fp
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 14 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     2920: contig of 2920 bp in length
        *     2921 3020: gap of      100 bp
        *     3021    32168: contig of 29148 bp in length
        *    32169 32268: gap of      100 bp
        *    32269    36080: contig of 3812 bp in length
        *    36081 36180: gap of      100 bp
        *    36181    45704: contig of 9524 bp in length
        *    45705 45804: gap of      100 bp
        *    45805    59680: contig of 13876 bp in length
        *    59681 59780: gap of      100 bp
        *    59781    61890: contig of 2110 bp in length
        *    61891 61990: gap of      100 bp
        *    61991    72185: contig of 10195 bp in length
        *    72186 72285: gap of      100 bp
        *    72286    76243: contig of 3958 bp in length
        *    76244 76343: gap of      100 bp
        *    76344    81840: contig of 5497 bp in length
        *    81841 81940: gap of      100 bp
        *    81941    86241: contig of 4301 bp in length
        *    86242 86341: gap of      100 bp
        *    86342   129124: contig of 42783 bp in length
        *   129125 129224: gap of      100 bp
        *   129225   131985: contig of 2761 bp in length
        *   131986 132085: gap of      100 bp
        *   132086   151439: contig of 19354 bp in length
        *   151440 151539: gap of      100 bp
        *   151540   168884: contig of 17345 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-569I12"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .2920
                     /note="assembly_fragment:00514
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3021. .32168
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    32269. .36080
                     /note="assembly_fragment:00777
                     fragment_chain:1"
     misc_feature    36181. .45704
                     /note="assembly_fragment:00971
                     fragment_chain:1"
     misc_feature    45805. .59680
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    59781. .61890
                     /note="assembly_fragment:02150
                     fragment_chain:2"
     misc_feature    61991. .72185
                     /note="assembly_fragment:02579
                     fragment_chain:2"
     misc_feature    72286. .76243
                     /note="assembly_fragment:00659
                     fragment_chain:2"
     misc_feature    76344. .81840
                     /note="assembly_fragment:00425
                     fragment_chain:2"
     misc_feature    81941. .86241
                     /note="assembly_fragment:02764
                     fragment_chain:2"
     misc_feature    86342. .129124
                     /note="assembly_fragment:00846
                     fragment_chain:2"
     misc_feature    129225. .131985
                     /note="assembly_fragment:01963"
     misc_feature    132086. .151439
                     /note="assembly_fragment:02120
                     fragment_chain:3"
     misc_feature    151540. .168884
                     /note="assembly_fragment:01758
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
BASE COUNT    46011 a  38577 c  39446 g  43544 t   1306 others
ORIGIN


alignment_scores:
            Quality:  343.00              Length:    340
              Ratio:  3.811                 Gaps:      2
 Percent Similarity:  26.471   Percent Identity: 25.882

alignment_block:
 US-09-215-435-179_COPY_24_121 x AL590400   ..

 Align seg 1/1  to: AL590400  from: 1  to: 168884

       8 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 24
         :::|||   ||||||||||||||||||||||||||||||||||||||||
  114919 TCCCTGCAGGAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGAC 114968

      24 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  114969 TGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGA 115018

      41 ysGlySerGluGlySerLeuCysGlnThrGlnVal..Phe..... 53
         ||||||||||||||||||||||||||||||||||  :::
  115019 AGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGTGGGTATCGCCGCCCGG 115068

      53 ..................................................... 53

  115069 GGGGAGCCAGAGGGGATCCAGGGGAAGTGGGAGCCAGGGAGGGCCCAGAT 115118

      53 ..................................................... 53

  115119 TCCTGGGGAGGAAAGAAGGGTAGGTGGGCGGAAATGCCCTGGAGCCTGGA 115168

      53 ..................................................... 53

  115169 ATTCCCCCGTGGGACTGCACCTCCAAGCCTTCTGGGGGAGCCTGTGGACG 115218

      53 ..................................................... 53

  115219 GAACCAATATGGTAGCCCACCCTAGGGTTTTGGAATCCTGGATTGGGGCG 115268

      53 ..................................................... 53

  115269 GGAAGGCGATGTTTCACTTTCTAATGTGGGTTCACATTCTCCAGGAGTCT 115318

      53 ..................................................... 53

  115319 TCTCAGCAAATTCCGAAAGTAGTACAGTAAATGGGCACAGCTAGGTCCAG 115368

      53 ..................................................... 53

  115369 CCTCTGCCCCAGAGCCCCGTCTTCATCCAGCTCAGGAGTTCTGCTGTGGC 115418
```

```
     53 ..................................................  53

 115419 CCCGGGTGTTCCTGGTGGAGGATGGGCAGAGGCAGCCTCTGAGCTGGGTG 115468

     53 ..................................................  53

 115469 TGTAGGGGAGGGAGTTGATTTGCCCATGGAAGGATTTAAAGACAAGACAT 115518

     53 ..................................................  53

 115519 AGAGGTCCTGCGATGGTGTGTCCATGGACTTAGGACAGGCAGATGGGTCA 115568

     53 ..................................................  53

 115569 GTGACAGGCAGGGTTACAAGGTCTGGAGCTGGGAGGACACTGAACAATCC 115618

     53 ..................................................  53

 115619 AGGAGCCAGCCTGGAGAAGGCAGCCAGAGGTGCCACTCTGAGCAGCTGGC 115668

     53 ..................................................  53

 115669 AAGCTATCATTTATCCCTGGATCCTGGCTGTGGGCAAGCACATGGGCCCT 115718

     53 ..................................................  53

 115719 GGAGCTGGGCTGGGCTGGGCTTAGGGGCTCAGGGAAGAGCTGCCGCCAAG 115768

     54 ...............................PheGlyGlnTyrArgA  59
                                      ||||||||||||||||
 115769 GTCGAGGTCAAAGTCCTGTCTTTCCCAGGTGTTCTTTGGCCAATATAGAG 115818

     59 laCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLys  75
        ||||||||||||||||||||||||||||||||||||||||||||||||||
 115819 CGTGTCCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAA 115868

     76 TrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLe  92
        ||||||||||||||||||||||||||||||||||||||||||||||||||
 115869 TGGCTTAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTT 115918

     92 uAlaLysLysMetPhePhe  98
        ||||||||||||||||||
 115919 GGCTAAGAAAATGTTCTTC 115937
```

```
seq_name: gb_om:CCCOLIP

seq_documentation_block:
LOCUS       CCCOLIP       452 bp    mRNA            MAM       31-MAR-1995
DEFINITION  Canine mRNA for colipase.
ACCESSION   X53564
VERSION     X53564.1  GI:841
KEYWORDS    colipase.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Fukuoka,S.I. and Scheele,G.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1990) Fukuoka S.-I., Scheele G.A., The Charles A.
            Dana Research Institute, Harvard Medical School, Cell and Molecular
            Biology, Dana 533, Beth Israel Hospital, 330 Brookline Ave.,
            Boston, MA 02215, USA
REFERENCE   2  (bases 1 to 452)
  AUTHORS   Fukuoka,S., Taniguchi,Y., Kitagawa,Y. and Scheele,G.
  TITLE     Full length cDNA sequence encoding canine pancreatic colipase
  JOURNAL   Nucleic Acids Res. 18 (18), 5549 (1990)
  MEDLINE   91016846
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /dev_stage="adult"
                     /tissue_type="pancreas"
                     /clone_lib="pUC-9"
                     /clone="pCL1"
     sig_peptide     1. .51
                     /product="signal peptide"
     CDS             1. .339
                     /codon_start=1
                     /product="precursor peptide"
                     /protein_id="CAA37636.1"
                     /db_xref="GI:842"
                     /db_xref="SWISS-PROT:P19090"
                     /translation="MEKILVLLLVALAVVYAVPDPRGIIIHLEDGELCLNSVQCKSKC
                     CHRATGLSLARCAPKASENSECSAKTLYGVYYKCPCERGLTCEGDKSIVGSITNTNFG
                     VCHDAGRSKK"
     mat_peptide     52. .336
                     /product="colipase"
     misc_feature    429. .435
                     /note="polyA signal"
     polyA_site      452
                     /note="polyA site"
BASE COUNT       90 a    146 c    117 g     99 t
ORIGIN


alignment_scores:
            Quality:  147.00             Length:     84
              Ratio:  2.625                Gaps:      1
 Percent Similarity:  66.667   Percent Identity: 36.905

alignment_block:
 US-09-215-435-179_COPY_24_121 x CCCOLIP   ..

 Align seg 1/1  to: CCCOLIP  from: 1  to: 452

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::::   :::|||::::::   ||
      82 CTGGAAGACGGCGAGCTCTGTCTGAACAGCGTCCAGTGCAAGAGCAAGTG 131

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         ||||:::||||||               :::::::||||||   |||:::|
     132 CTGCCACCGGGCCACCGGGCTGAGCCTGGCCCGCTGCGCACCCAAGGCCA 181

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::::|||   |||
     182 GCGAGAACAGCGAGTGCTCTGCCAAGACGCTCTATGGGGTTTACTACAAG 231

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn......Gl 74
         |||||||||   |||:::|||||||||      :::|||:::
     232 TGTCCCTGTGAGCGGGGCCTGACCTGTGAGGGTGACAAGTCCATCGTGGG 281

      74 uLysTrpLeuSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnL 91
          :::      :::      :::|||   |||:::      |||||||:::|
     282 CTCCATCACCAATACCAACTTTGGTGTCTGCCACGATGCTGGGCGCTCCA 331

      91 ys 91
         ||
     332 AG 333
```

```
seq_name: gb_om:RABCOLIPAS

seq_documentation_block:
LOCUS       RABCOLIPAS    459 bp    mRNA            MAM       05-OCT-1993
DEFINITION  Oryctolagus cuniculus pancreatic colipase mRNA, complete cds.
ACCESSION   L06329
VERSION     L06329.1  GI:164893
KEYWORDS    colipase.
SOURCE      Oryctolagus cuniculus male adult pancreas cDNA to mRNA.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Colwell,N.S., Aleman-Gomez,J.A., Sasser,T. and Kumar,V.B.
```

```
  TITLE     Cloning and characterization of rabbit pancreatic colipase
  JOURNAL   Int. J. Biochem. 25 (6), 885-890 (1993)
  MEDLINE   93345715
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
     sig_peptide     39. .89
     CDS             39. .362
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA02911.1"
                     /db_xref="GI:164894"
                     /translation="MEKVLVLLLVALSVAYAAPGPRGIVINLEEGELCLNSAQCKSGC
                     CHHSSALSLARCAPKASENSECSPQTIYGVYYKCPCERGLTCEGDKSIVGSITNTNFG
                     VCLDV"
     mat_peptide     90. .359
                     /product="colipase"
BASE COUNT       84 a    170 c    122 g     83 t
ORIGIN


alignment_scores:
            Quality:  141.00              Length:     59
              Ratio:  3.279                 Gaps:      0
 Percent Similarity:  72.881   Percent Identity: 42.373

alignment_block:
 US-09-215-435-179_COPY_24_121 x RABCOLIPAS   ..

 Align seg 1/1  to: RABCOLIPAS  from: 1  to: 459

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::|||||
     120 CTGGAGGAGGGCGAGCTCTGCTTGAACAGCGCCCAGTGCAAGAGCGGCTG 169

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         ||||:::::::::                :::::|||||   |||:::|
     170 CTGCCACCACTCGAGCGCTCTGAGTCTGGCCCGCTGCGCACCCAAGGCCA 219

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::   |||:::::::::|||   |||
     220 GCGAGAACAGCGAGTGTTCGCCCCAGACCATCTACGGCGTCTACTACAAG 269

      60 CysProCysLeuArgAsnLeuThrCys 68
         |||||||||   |||:::|||||||||
     270 TGTCCCTGTGAGCGCGGCCTGACCTGC 296

 seq_name: gb_pr10:HUMCOLIP

 seq_documentation_block:
 LOCUS       HUMCOLIP      523 bp    mRNA            PRI       01-NOV-1994
 DEFINITION  Human colipase mRNA, complete cds.
 ACCESSION   J02883
 VERSION     J02883.1  GI:180885
 KEYWORDS    cofactor; colipase; triglyceride lipase.
 SOURCE      Human adult pancreas, cDNA to mRNA.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 523)
   AUTHORS   Lowe,M.E., Rosenblum,J.L., McEwen,P. and Strauss,A.W.
   TITLE     Cloning and characterization of the human colipase cDNA
   JOURNAL   Biochemistry 29 (3), 823-828 (1990)
   MEDLINE   90248429
 COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
             by M.E.Lowe, 17-NOV-1989.
 FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6pter-p21.1"
     mRNA            <1. .523
                     /note="CLP mRNA"
     gene            22. .360
                     /gene="CLPS"
     CDS             22. .360
                     /gene="CLPS"
                     /note="colipase precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-127-277"
                     /protein_id="AAA52054.1"
                     /db_xref="GI:180886"
                     /translation="MEKILILLLVALSVAYAAPGPRGIIINLENGELCMNSAQCKSNC
                     CQHSSALGLARCTSMASENSECSVKTLYGIYYKCPCERGLTCEGDKTIVGSITNTNFG
                     ICHDAGRSKQ"
     sig_peptide     23. .73
                     /gene="CLPS"
                     /note="colipase signal peptide"
     mat_peptide     74. .357
                     /gene="CLPS"
                     /note="colipase"
BASE COUNT      109 a    173 c    128 g    113 t
ORIGIN


alignment_scores:
            Quality:  135.00              Length:     87
              Ratio:  2.328                 Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x HUMCOLIP   ..

 Align seg 1/1  to: HUMCOLIP  from: 1  to: 523

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::||
     103 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 152

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                :::::|||::::::   :::|
     153 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 202

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::::|||   |||
     203 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 252

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||          :::::::::|||
     253 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 293

      76 pLeu.............SerIleAlaTyrGlyArgCysGlnLysIleG 88
         :::                :::      :::|||   |||:::     |
     294 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 343

      88 lyArgGlnLys 91
         |||||:::|||
     344 GACGCTCCAAG 354

 seq_name: gb_pat1:AX014220

 seq_documentation_block:
 LOCUS       AX014220      656 bp    DNA             PAT       07-SEP-2000
 DEFINITION  Sequence 16 from Patent WO9954446.
 ACCESSION   AX014220
 VERSION     AX014220.1  GI:10040621
 KEYWORDS    .
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 656)
   AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
             Pilarsky,C.
   TITLE     Human nucleic acid sequences from normal pancreas tissue
```

```
  JOURNAL   Patent: WO 9954446-A 16 28-OCT-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      165 a    200 c    162 g    129 t
ORIGIN


alignment_scores:
            Quality:  135.00             Length:     87
              Ratio:  2.328                Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x AX014220   ..

 Align seg 1/1  to: AX014220  from: 1  to: 656

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| |
     170 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 219

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                ::::::|||::::::   :::|
     220 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 269

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::|||   |||
     270 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 319

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         :::::::::|||
     320 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 360

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::             :::      :::|||   |||:::     |
     361 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 410

      88 lyArgGlnLys 91
         |||||:::|||
     411 GACGCTCCAAG 421

seq_name: gb_rol:MCCOLIPAS

seq_documentation_block:
LOCUS       MCCOLIPAS     490 bp    mRNA            ROD       21-NOV-1996
DEFINITION  M.coypus mRNA for colipase.
ACCESSION   X82998
VERSION     X82998.1  GI:599866
KEYWORDS    colipase.
SOURCE      nutria.
  ORGANISM  Myocastor coypus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Hystricognathi; Myocastoridae;
            Myocastor.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Thirstrup,K., Carriere,F., Hjorth,S.A., Rasmussen,P.B.,
            Nielsen,P.F., Ladefoged,C., Thim,L. and Boel,E.
  TITLE     Cloning and expression in insect cells of two pancreatic lipases
            and a procolipase from Myocastor coypus
  JOURNAL   Eur. J. Biochem. 227 (1-2), 186-193 (1995)
  MEDLINE   95154288
REFERENCE   2  (bases 1 to 490)
  AUTHORS   Thirstrup,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1994) K. Thirstrup, Univ. Hospital, Lab. for
            Molecular Endocrinology, Rigshospitalet 6321, Blegdamsvej 9,
            DK-2100 Copenhagen, DENMARK
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Myocastor coypus"
                     /db_xref="taxon:10157"
                     /tissue_type="pancreas"
     CDS             46. .381
                     /codon_start=1
                     /product="colipase"
                     /protein_id="CAA58119.1"
                     /db_xref="GI:599867"
                     /db_xref="SWISS-PROT:P42889"
                     /translation="MEKVLALVLLTLAVAYAAPDPRGLIINLDNGELCLNSAQCKSQC
                     CQHDSPLGLARCADKARENSGCSPQTIYGIYYLCPCERGLTCDGDKSIIGAITNTNYG
                     ICQDPQSKK"
     sig_peptide     46. .96
     mat_peptide     97. .378
                     /note="putative propeptide: 97. .111"
                     /product="colipase"
     misc_feature    97. .111
                     /note="putative propeptide"
                     /product="colipase"
BASE COUNT       96 a    181 c    120 g     93 t
ORIGIN


alignment_scores:
            Quality:  134.50             Length:     95
              Ratio:  2.360                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 34.737

alignment_block:
 US-09-215-435-179_COPY_24_121 x MCCOLIPAS   ..

 Align seg 1/1  to: MCCOLIPAS  from: 1  to: 490

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         |||:::      |||   |||::::::::::   :::|||::::::   ||
     127 CTGGACAACGGCGAGCTCTGCCTGAACAGCGCCCAGTGCAAGAGTCAGTG 176

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||:::                  ::::::||||||:::|||:::
     177 CTGCCAGCACGACAGCCCCCTGGGCCTGGCCCGCTGCGCCGACAAGGCCA 226

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
           |||:::|||   |||:::   |||::::::::::|||   |||
     227 GAGAGAACAGCGGCTGCTCCCCACAGACGATCTATGGAATCTACTACTTG 276

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn......Gl 74
         |||||||||   |||:::|||||||||      :::|||:::
     277 TGTCCCTGCGAGCGGGGCCTGACCTGTGACGGGGACAAGTCCATCATCGG 326

      74 uLysTrpLeuSerIleAlaTyrGlyArgCysGln..............L 86
                :::      ||||||   ||||||               |
     327 CGCCATCACCAACACCAACTACGGCATCTGCCAGGATCCTCAGTCCAAGA 376

      86 ysIleGlyArgGlnLysLeuAlaLysLysMetPhe 97
         ||   |||::::::   ||||||   :::   |||
     377 AGTGAGGCCACCACGCGCTGGCCCTGTCCCCTTTC 411

seq_name: gb_om:ECCOLIPB

seq_documentation_block:
LOCUS       ECCOLIPB      497 bp    mRNA            MAM       11-NOV-1994
DEFINITION  E.caballus mRNA for colipase B.
ACCESSION   X74344
VERSION     X74344.1  GI:572678
KEYWORDS    colipase.
SOURCE      horse.
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Kerfelec,B.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (28-JUL-1993) B. Kerfelec, CNRS, CBM, 31 Chemin Joseph
            Aiguier, 13402 Marseille, Cedex 9, FRANCE
REFERENCE   2  (bases 1 to 497)
  AUTHORS   Crenon,I., Granon,S., Chapus,C. and Kerfelec,B.
  TITLE     Molecular cloning and expression of two horse pancreatic cDNA
            encoding colipase A and B
  JOURNAL   Biochim. Biophys. Acta 1213 (3), 357-360 (1994)
  MEDLINE   94325330
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /tissue_type="pancreas"
     CDS             <1. .327
                     /codon_start=1
                     /product="colipase B"
                     /protein_id="CAA52391.1"
                     /db_xref="GI:572679"
                     /db_xref="SWISS-PROT:P02705"
                     /translation="LALLLVALAVAYAVPDPRGVIINLEAGEICMNSAQCKSECCHRE
                     SSLSLARCAAKASENSECSAWTLYGVYYKCPCERGLTCQVDKTLVGSIMNTNFGICFD
                     AARSEE"
BASE COUNT      118 a    161 c    125 g     93 t
ORIGIN


alignment_scores:
            Quality:  134.00            Length:     63
              Ratio:  3.045               Gaps:      0
 Percent Similarity:  69.841   Percent Identity: 39.683

alignment_block:
 US-09-215-435-179_COPY_24_121 x ECCOLIPB   ..

 Align seg 1/1  to: ECCOLIPB  from: 1  to: 497

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||::::::   ||
      70 TTGGAGGCGGGCGAGATCTGCATGAACAGCGCCCAGTGCAAGAGCGAATG 119

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         ||||:::|||      :::         ::::::||||||   |||:::|
     120 CTGCCACCGGGAAAGCAGCCTGAGCCTGGCCCGCTGCGCAGCCAAAGCCA 169

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||::::::   ::::::::::|||   |||
     170 GCGAGAACAGCGAGTGCTCTGCCTGGACACTCTACGGGGTTTACTACAAG 219

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLys 72
         |||||||||   |||:::|||||||||      :::|||
     220 TGTCCCTGTGAGCGGGGCCTGACCTGCCAGGTGGACAAG 258

 seq_name: gb_om:AF148567

 seq_documentation_block:
 LOCUS       AF148567      510 bp    mRNA            MAM       06-MAY-2000
 DEFINITION  Sus scrofa pancreatic colipase mRNA, complete cds.
 ACCESSION   AF148567
 VERSION     AF148567.1  GI:7711135
 KEYWORDS    .
 SOURCE      pig.
   ORGANISM  Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE   1  (bases 1 to 510)
   AUTHORS   Darnis,S. and Chaix,J.C.
   TITLE     Cloning, sequencing and functional expression of porcine pancreatic
             colipase
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 510)
   AUTHORS   Darnis,S. and Chaix,J.C.
   TITLE     Direct Submission
   JOURNAL   Submitted (04-MAY-1999) LBBN, CNRS.ESA6033, Av Escadrille Normandie
             Niemen, Marseille 13397, France
 FEATURES             Location/Qualifiers
      source          1. .510
                      /organism="Sus scrofa"
                      /db_xref="taxon:9823"
                      /tissue_type="pancreas"
      CDS             44. .382
                      /note="cofactor; triglyceride lipase"
                      /codon_start=1
                      /product="colipase"
                      /protein_id="AAF67823.1"
                      /db_xref="GI:7711136"
                      /translation="MEKVLALLLVTLTVAYAVPDPRGIIINLDEGELCLNSAQCKSNC
                      CQHDTILSLSRCALKARENSECSAFTLYGVYYKCPCERGLTCEGDKSLVGSITNTNFG
                      ICHDVGRSSD"
 BASE COUNT      117 a    160 c    121 g    112 t
 ORIGIN


 alignment_scores:
             Quality:  133.00            Length:     85
               Ratio:  2.463               Gaps:      2
  Percent Similarity:  63.529   Percent Identity: 34.118

 alignment_block:
  US-09-215-435-179_COPY_24_121 x AF148567   ..

  Align seg 1/1  to: AF148567  from: 1  to: 510

       10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
          |||:::      |||   |||:::::::::   :::|||::::::::||
      125 CTGGATGAGGGCGAGCTCTGCCTGAACAGTGCCCAGTGCAAGAGCAACTG 174

       26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
          |||||||:::                  |||:::||||||   |||:::
      175 CTGCCAGCATGACACAATCCTGAGCCTGTCCCGCTGCGCACTCAAGGCCA 224

       43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
          |||:::|||   |||::::::   :::::::::|||   |||
      225 GAGAGAACAGCGAGTGTTCTGCCTTCACGCTCTATGGGGTTTACTACAAG 274

       60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
          |||||||||   |||:::|||||||||         :::::::|||
      275 TGTCCCTGTGAACGGGGCCTGACCTGT.........GAGGGGACAAGAG 315

       76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          |||               :::      :::|||   |||:::   :::|
      316 TCTCGTGGGCTCCATCACCAACACCAACTTTGGTATCTGCCATGATGTTG 365

       88 lyArg 89
          |||||
      366 GACGC 370

 seq_name: gb_om:ECCO1IPA

 seq_documentation_block:
 LOCUS       ECCO1IPA      472 bp    mRNA            MAM       28-JUL-1995
 DEFINITION  E.caballus mRNA for colipase A.
 ACCESSION   X74503
 VERSION     X74503.1  GI:572676
 KEYWORDS    anchoring protein; colipase A; secreted protein; triacylglycerol
             digestion.
 SOURCE      horse.
   ORGANISM  Equus caballus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 REFERENCE   1  (bases 1 to 472)
   AUTHORS   Kerfelec,B.
   TITLE     Direct Submission
   JOURNAL   Submitted (09-AUG-1993) B. Kerfelec, CNRS, CBM, 31 Chemin Joseph
             Aiguier, 13402 Marseille, Cedex 9, FRANCE
 REFERENCE   2  (bases 1 to 472)
   AUTHORS   Crenon,I., Granon,S., Chapus,C. and Kerfelec,B.
```

```
  TITLE     Molecular cloning and expression of two horse pancreatic cDNA
            encoding colipase A and B
  JOURNAL   Biochim. Biophys. Acta 1213 (3), 357-360 (1994)
  MEDLINE   94325330
COMMENT     conflicts with previous reported amino acid sequence (Pierrot et
            al., 1982, Eur.J.Biochem. 123, 347-354:E22->Q22; E32->S32;
            N89->D89; K92->R92; S/E93->S93; R94->E94; E95.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /clone_lib="cDNA library in lambda gt10"
                     /tissue_type="pancreas"
     sig_peptide     <1. .33
     CDS             <1. .321
                     /codon_start=1
                     /product="colipase A precursor"
                     /protein_id="CAA52611.1"
                     /db_xref="GI:572677"
                     /db_xref="SWISS-PROT:P02704"
                     /translation="LLLVALAVAYAVPDPRGVIINLEAGEICLNSAQCKSECCHQESS
                     LSLARCAAKASENSECSAWTLYGVYYKCPCERGLTCQVDKTLVGSIMNTNFGICFDAA
                     RSEE"
     mat_peptide     34. .318
                     /gene="colipase A"
                     /product="colipase A"
     gene            34. .318
                     /gene="colipase A"
     misc_feature    196. .210
                     /gene="colipase A"
                     /note="interfacial recognition site"
     3'UTR           319. .471
     polyA_signal    441. .447
BASE COUNT       99 a    158 c    123 g     92 t
ORIGIN


alignment_scores:
            Quality:  131.00             Length:     63
              Ratio:  2.977                Gaps:      0
 Percent Similarity:  69.841   Percent Identity: 38.095

alignment_block:
 US-09-215-435-179_COPY_24_121 x ECCO1IPA   ..

 Align seg 1/1  to: ECCO1IPA  from: 1  to: 472

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||::::::   ||
      64 TTGGAGGCGGGCGAGATCTGCCTGAACAGCGCCCAGTGCAAGAGCGAATG 113

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         ||||::::::      :::         ::::::||||||   |||:::|
     114 CTGCCACCAGGAAAGCAGCCTGAGCCTGGCCCGCTGTGCAGCCAAAGCCA 163

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||::::::   :::::::::|||   |||
     164 GCGAGAACAGCGAGTGCTCTGCCTGGACACTCTACGGGGTTTACTACAAG 213

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLys 72
         |||||||||   |||:::||||||||||      :::|||
     214 TGTCCCTGTGAGCGGGGCCTGACCTGCCAGGTGGACAAG 252

seq_name: gb_ro2:RATCOLIP

seq_documentation_block:
LOCUS       RATCOLIP      446 bp    mRNA            ROD       27-APR-1993
DEFINITION  Rat colipase mRNA, complete cds.
ACCESSION   M33333
VERSION     M33333.1  GI:203502
KEYWORDS    colipase.
SOURCE      Rattus norvegicus (strain Wistar) (library: puc) pancreas cDNA to
            mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Wicker,C. and Puigserver,A.
  TITLE     Rat pancreatic colipase mRNA: Nucleotide sequence of a cDNA clone
            and nutritional regulation by a lipidic diet
  JOURNAL   Biochem. Biophys. Res. Commun. 167, 130-136 (1990)
  MEDLINE   90179738
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /tissue_type="pancreas"
                     /tissue_lib="puc"
     mRNA            <1. .446
                     /note="colipase mRNA"
     CDS             18. .356
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA40943.1"
                     /db_xref="GI:203503"
                     /translation="MKVLVVLLVTLVAVAYAVPGPRGLFINLEDGEICVNSMQCKSRC
                     CQHDTILGIARCTHKAMENSECSPKTLYGIYYRCPCERGLTCEGDRSIIGAITNTNYG
                     VCLDSTRSKQ"
     sig_peptide     18. .68
     mat_peptide     69. .353
                     /product="colipase"
     polyA_signal    424. .430
     polyA_site      446
BASE COUNT       96 a    141 c    113 g     96 t
ORIGIN


alignment_scores:
            Quality:  122.00             Length:     91
              Ratio:  2.392                Gaps:      2
 Percent Similarity:  56.044   Percent Identity: 31.868

alignment_block:
 US-09-215-435-179_COPY_24_121 x RATCOLIP   ..

 Align seg 1/1  to: RATCOLIP  from: 1  to: 446

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::::::::::|||::::::   ||
      99 CTGGAGGACGGTGAGATCTGCGTAAACAGTATGCAGTGTAAGAGCAGATG 148

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||:::                  :::::::|||:::::::|||:::
     149 CTGCCAACATGACACCATCCTGGGCATCGCCCGGTGCACACACAAGGCCA 198

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
           |||:::|||   |||:::   :::::::::::::|||   |||
     199 TGGAGAACAGCGAGTGCTCCCCAAAGACCCTCTATGGGATCTACTACAGG 248

      60 CysProCysLeuArgAsnLeuThrCys..................... 68
         |||||||||   |||:::|||||||||
     249 TGTCCCTGTGAGCGGGGCCTGACCTGTGAGGGGGACAGGAGCATCATTGG 298

      69 ....IleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
             |||   :::   |||                      ||||||   |
     299 CGCCATCACCAACACCAAC.....................TACGGCGTCT 327

      84 ysGlnLysIleGlyArgGlnLys 91
         ||           |||:::|||
     328 GCCTCGACTCCACCCGCTCCAAG 350

seq_name: gb_ro2:RATCOLQ

seq_documentation_block:
```

```
LOCUS       RATCOLQ       492 bp    mRNA            ROD       04-AUG-1994
DEFINITION  Rat colipase mRNA, complete cds.
ACCESSION   M58370
VERSION     M58370.1  GI:203504
KEYWORDS    colipase.
SOURCE      Rat, cDNA to mRNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Payne,R.M., Sims,H.F., Jennens,M.L. and Lowe,M.E.
  TITLE     Rat pancreatic lipase and two related proteins: enzymatic
            properties and mRNA expression during development
  JOURNAL   Am. J. Physiol. 266 (5 Pt 1), G914-G921 (1994)
  MEDLINE   94262798
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             59. .397
                     /codon_start=1
                     /product="colipase"
                     /protein_id="AAA20505.1"
                     /db_xref="GI:203505"
                     /translation="MKVLVVLLVTLVAVAYAAPGPRGLFINLEDGEICVNSMQCKSRC
                     CQHDTILGIARCTHKAMENSECSPKTLYGIYYRCPCERGLTCEGDRSIIGAITNTNYG
                     VCLDSTRSKQ"
BASE COUNT      107 a    163 c    119 g    103 t
ORIGIN


alignment_scores:
            Quality:  122.00            Length:     91
              Ratio:  2.392               Gaps:      2
 Percent Similarity:  56.044   Percent Identity: 31.868

alignment_block:
 US-09-215-435-179_COPY_24_121 x RATCOLQ   ..

 Align seg 1/1  to: RATCOLQ  from: 1  to: 492

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||     |||   |||:::::::::::::::::|||::::::   ||
     140 CTGGAGGACGGTGAGATCTGCGTAAACAGTATGCAGTGTAAGAGCAGATG 189

       26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
          |||||||:::                 :::::::|||:::::::|||:::
     190 CTGCCAACATGACACCATCCTGGGCATCGCCCGGTGCACACACAAGGCCA 239

       43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
            |||:::|||   |||:::   :::::::::::::|||   |||
     240 TGGAGAACAGCGAGTGCTCCCCAAAGACCCTCTATGGGATCTACTACAGG 289

       60 CysProCysLeuArgAsnLeuThrCys..................... 68
          |||||||||   |||:::||||||||||
     290 TGTCCCTGTGAGCGGGGCCTGACCTGTGAGGGGGACAGGAGCATCATTGG 339

       69 ....IleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
             |||   :::   |||                    ||||||   |
     340 CGCCATCACCAACACCAAC.....................TACGGCGTCT 368

       84 ysGlnLysIleGlyArgGlnLys 91
          ||          |||:::|||
     369 GCCTCGACTCCACCCGCTCCAAG 391

 seq_name: gb_in3:CEW01B6

 seq_documentation_block:
 LOCUS       CEW01B6     28122 bp    DNA             INV       27-OCT-2000
 DEFINITION  Caenorhabditis elegans cosmid W01B6, complete sequence.
 ACCESSION   Z68301
 VERSION     Z68301.1  GI:1130644
 KEYWORDS    HTG; Collagen; Protein kinase; Tyrosine phosphatase;
             Tyrosine-protein kinase; WNT-2.
 SOURCE      Caenorhabditis elegans.
   ORGANISM  Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE   1  (bases 1 to 28122)
   AUTHORS   none.
   TITLE     Genome sequence of the nematode C. elegans: a platform for
             investigating biology. The C. elegans Sequencing Consortium
   JOURNAL   Science 282 (5396), 2012-2018 (1998)
   MEDLINE   99069613
   REMARK    The C.elegans Sequencing Consortium.
             Erratum:[[published errata appear in Science 1999 Jan
             1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
             3;285(5433):1493]]
 REFERENCE   2  (bases 1 to 28122)
   AUTHORS   Coles,L.
   TITLE     Direct Submission
   JOURNAL   Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Centre,
             Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
             Washington University, St. Louis, MO 63110, USA. E-mail:
             jes@sanger.ac.uk or rw@nematode.wustl.edu
 COMMENT     Coding sequences below are predicted from computer analysis, using
             predictions from Genefinder (P. Green, U. Washington), and other
             available information.
             Current sequence finishing criteria for the C. elegans genome
             sequencing consortium are that all bases are either sequenced
             unambiguously on both strands, or on a single strand with both a
             dye primer and dye terminator reaction, from distinct subclones.
             Exceptions are indicated by an explicit note.
             IMPORTANT: This sequence is not the entire insert of clone W01B6.
             It may be shorter because we only sequence overlapping sections
             once, or longer because we arrange for a small overlap between
             neighbouring submissions.
             The true left end of clone W01B6 is at 1 in this sequence. The true
             right end of clone W01B6 is at 7269 in
             sequence Z70718.
             The true left end of clone C04G2 is at 28019 in this sequence. The
             true right end of clone T04B2 is at 16662 in this sequence. The
             start of this sequence (1. .104) overlaps with the end of sequence
             Z68299.
             The end of this sequence (28019. .28122) overlaps with the start of
             sequence Z70718.
             For a graphical representation of this sequence and its analysis
             see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
             name=W01B6.
 FEATURES             Location/Qualifiers
      source          1. .28122
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="IV"
                      /clone="W01B6"
      gene            join(9006. .9085,9340. .9493,9542. .9927,9972. .10107,
                      10154. .10314,10932. .11097)
                      /gene="cwn-2"
                      /note="W01B6.1"
      CDS             join(9006. .9085,9340. .9493,9542. .9927,9972. .10107,
                      10154. .10314,10932. .11097)
                      /gene="cwn-2"
                      /note="contains similarity to Pfam domain: PF00110 (wnt
                      family of developmental signaling proteins), Score=700.5,
                      E-value=6.2e-255, N=1
                      cDNA EST EMBL:Z14953 comes from this gene
                      cDNA EST yk343h8.3 comes from this gene
                      cDNA EST yk343h8.5 comes from this gene"
                      /codon_start=1
                      /protein_id="CAA92624.1"
                      /db_xref="GI:3880389"
                      /db_xref="SWISS-PROT:P34889"
                      /translation="MIPRRSCWLILLLNLLNVQSLLDASWWSTVAQLSTALAGHNVKP
                      VCELPGLSPGQAQVCELFKDHMPAVSIGAQNAIQECQRQFTGHRWNCSTHYSTGMLGP
                      IHKMATREAAFTYAILSAGVTHEIGRRCKQGLLTSCGCSDETKPKNVPTDWSWGGCGD
```

```
                NVEYGYKFSRDFIDIREKEHDPKRNHDNGRSLMNRRNNEAGRKILKRHRKPKCKCHGV
                SGACNMKTCWMQLPSMEQVGKILRNKYDKAIRVQINDRGNLQLLADEATKERKTRALP
                TDLVFMDDSPDYCRFDRHSGTLGTEGRVCKRGSGGAEGCDSLCCGRGYNTYTQEVKSK
                CNCKFEWCCKVVCQTCNNVTQVDICK"
     gene       complement(join(11401. .11572,11702. .11769,11814. .12098,
                12646. .13053,13098. .13463,13533. .13574))
                /gene="W01B6.3"
     CDS        complement(join(11401. .11572,11702. .11769,11814. .12098,
                12646. .13053,13098. .13463,13533. .13574))
                /gene="W01B6.3"
                /note="Similarity to E.coli tetracycline resistance
                protein (SW:TCR4_ECOLI)"
                /codon_start=1
                /protein_id="CAA92626.1"
                /db_xref="GI:3880391"
                /db_xref="SPTREMBL:Q23104"
                /translation="MSIYFTSTWQYLSENDETATVDFFGAILALQALASALTNPLFGY
                WNQVSSSTRKPIFCACGISAIGNVLYAFAYLAPPFVKWVMLFARFMTGFSPGALGVLR
                SFIGTASTQEDRMKAVSITNAGFTAGFFLGPTIQICFIPLGKDGFIFGSFVLNMYTSA
                AVFMSIASIFSILLTYFYLEENYVGIISKEDKTSDPYFVLPKFDRLPVLLLFYMWWLM
                CGVVCVESMAAPVTIAMFNWSREDAVLYNGIIQTISCIFTFIVNFSLASTRLKTIDNR
                LLLLCGLIFFEVFFIFHMPLPFYPGPLDRPDTMNSTTTLIGRCDFDWCDDTPRVPLPL
                YLFISSIVIGLGFPLISSTSSSLLSQIIGPRKQGTVQGFFAFTGSFSQFVVSLFSTRL
                FELSGYKWIMMYHWCIVTVAVFSVILLWRRLVPLKITPTNGEATRYKLGTFYRM"
     gene       complement(join(14566. .14703,14747. .14834,14876. .15195,
                15241. .15588,15968. .16115,16164. .16222))
                /gene="W01B6.2"
     CDS        complement(join(14566. .14703,14747. .14834,14876. .15195,
                15241. .15588,15968. .16115,16164. .16222))
                /gene="W01B6.2"
                /note="contains similarity to Pfam domain: PF00069
                (Eukaryotic protein kinase domain), Score=85.0,
                E-value=5.5e-23, N=1
                cDNA EST EMBL:AW057393 comes from this gene
                cDNA EST EMBL:AW057450 comes from this gene
                cDNA EST EMBL:AW057476 comes from this gene"
                /codon_start=1
                /protein_id="CAA92625.1"
                /db_xref="GI:3880390"
                /db_xref="SPTREMBL:Q23103"
                /translation="MPSEKEDDVMKNVTFAEGKKFGDWKIGKTIDEGGFGKVYIATSI
                SDPKKVAALKAESNEIEGGSAIKLEAMILNKLNANGPVPHIPVVHLCAKRKLYCYMVM
                TLLGRNLRKLKSTNLVVNNGFSRGTWSRIGIQCLYALKYVHDNGFIHRDVKPQNFLLG
                NETDSERARIVHILDFGLARPFAVFHARENKWIARRARGTAEFRGTLRYTSPNVHLRK
                EQGRVDDVWSLLYVIIELNGGKALPWQTDSQRERVEQMKLNLPAKVVMSNMPACMDKV
                MPHLASLDYYQRPDYHMIFKCFWQVMENEKITPSSKYDWEKDEPDKSVPPASWENPAG
                PFFQSDPLGINAPPTPAEAALTEKSGKKHGKK"
     gene       16591. .16782
                /gene="W01B6.4"
     CDS        16591. .16782
                /gene="W01B6.4"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA92622.1"
                /db_xref="GI:3880387"
                /db_xref="SPTREMBL:Q23101"
                /translation="MCLVQATGGSSTHKLVSAGAEKMIITRPDFVDPPGSKDISITRT
                ARAPKEDKLVIHFCTCRCN"
     gene       join(17456. .17665,17912. .18275,18438. .18601,18645. .18966,
                19023. .19175,19227. .19510,19569. .19682,19730. .19780)
                /gene="W01B6.5"
     CDS        join(17456. .17665,17912. .18275,18438. .18601,18645. .18966,
                19023. .19175,19227. .19510,19569. .19682,19730. .19780)
                /gene="W01B6.5"
                /note="predicted using Genefinder
                Similarity to Mouse tyrosine-protein kinase
                (SW:LYN_MOUSE), contains similarity to Pfam domain:
                PF00069 (Eukaryotic protein kinase domain), Score=164.4,
                E-value=6.1e-46, N=1"
                /codon_start=1
                /protein_id="CAA92623.1"
                /db_xref="GI:3880388"
                /db_xref="SPTREMBL:Q23102"
                /translation="MCTQDEECSAPGGDKGKEKEEHTVVPRGKEEKEKKPGTKPSPIS
                PAFKGTPKERPNCGKTKNETSGKKLRSNDVSPDVKDRLQLSEHGPSKTCIKDFQWLED
                FITKMPCFHGFIGREDLMSVLKNVGDFLIRTSVQPNKHEVEKMKKNQADVKVLGNLSR
                EKCAKKEKEKEEKLAGMGDVGRREFVISVYCKDKEKPATDKPQYTPIRNLVIKRDNGM
                VHVEPLRKFKTLTEFFGYYQKNSGICKETDFQLITPIPLSNWEFIHEDIALQQKKLGE
                GAFGEVRIGKMKLKSTKKTVEVAVKMLRNAEVVIREQVGELLHEARVMRMMDHKNVLR
                SYGIAVLKEPLYLMTEMCACGALREYLRENQETVTLAEKLFFVVGSSRGVQYLHSQKT
                IHRDLAVRNILLIHKRHDPAITNSTVSMGTFLFQPKISDFGLAKISERYEMKEQCKIP
                VRYLAPETLESFIFTTKTDVFSFGCVIWEIYENGNQPHDGKNAQTIRNLTKKNQFLKL
                TNSAPSELRKLIEERVFTSDPENRCSMTTIVQCAEKIEKPPVGMK"
     gene       join(22357. .22615,22773. .22963,23009. .23092,23136. .23432,
                23478. .23732,23904. .23965,24007. .24100)
                /gene="W01B6.6"
     CDS        join(22357. .22615,22773. .22963,23009. .23092,23136. .23432,
                23478. .23732,23904. .23965,24007. .24100)
                /gene="W01B6.6"
                /note="contains similarity to Pfam domain: PF00102
                (Protein-tyrosine phosphatase), Score=172.6,
                E-value=2e-48, N=1"
                /codon_start=1
                /protein_id="CAA92627.1"
                /db_xref="GI:3880392"
                /db_xref="SPTREMBL:Q23105"
                /translation="MSKSTRSRRKKKSKGGDRSERKTNKSERKKKGAKSDIGKKKKKN
                ETVSKDQTNTGVSTGAEGVEKKEKEQWSGDETAKKLVASGAFNTSTITSGFKELSNPK
                PSADMCSIFKNNPKKSRAPDWPIADNKLIKLSHAPGNFICAAKVDVPEFNRTMIVTQI
                PDVTVSANIEDFWRMIFQEEIVSIVIAVMPLECSVTLQQLFPLLSGTFSNHGKMFLNN
                KKVDSAVAMTAYTLEVLPDGCSNSLLSTVYHLHNWKQKQGLENVGELVTTVEKVLKTN
                ENTVLMSMNGIGRAGTMLTLFTSMLSVQKSKEVNPKEILIKLRGERSGLVENADQFNT
                VHRAMALWFKNKCSDEEVQKKVVEFAPNLTTEAVERSSYEYLNSQHFLYPILLESMLH
                QHQMKQVLLKKVERSMERSKL"
     gene       complement(join(24622. .24664,24714. .24833,24928. .25089,
                25317. .25504))
                /gene="W01B6.8"
     CDS        complement(join(24622. .24664,24714. .24833,24928. .25089,
                25317. .25504))
                /gene="W01B6.8"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA92621.1"
                /db_xref="GI:3880386"
                /db_xref="SPTREMBL:Q23100"
                /translation="MFNNFPFSKFPFISQYFNITSEGCSSKIMCLYSQISIILFLFWS
                LMLQLIIGTTCVSKNSQKRIDSLKVPRRAPQAQRFTNSRISSNNQVSLRNSQKSETES
                AKTEIINVPLGKAAKDETSSNILTPANKQKTDLHTAPTRKAQFADTLKRVKRRATKPD
                DPDEKTSYIP"
     gene       complement(join(25616. .26241,26289. .26568))
                /gene="col-2"
                /note="W01B6.7"
     CDS        complement(join(25616. .26241,26289. .26568))

alignment_scores:
           Quality:   90.00            Length:     75
             Ratio:  1.957              Gaps:      5
 Percent Similarity:  61.333   Percent Identity: 34.667

alignment_block:
 US-09-215-435-179_COPY_24_121 x CEW01B6   ..

 Align seg 1/1  to: CEW01B6  from: 1  to: 28122

      14 GluSerCysIleArgAsnGlnAspCysGlu.................Th 24
         ||||||||||||||||||   :::|||                    ||
   13131 GAATCCTGCATTCGTAATTGAAACTGCTTTCATCCGATCTTCTTGTGTAC 13180

      24 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 41
         |   |||   :::::::::   :::|||   ||||||      :::|||
   13181 TTGCTGTACCGATAAAACTTCTGAGAACTCCGAGAGCACCAGGACTGAAA 13230

      41 ysGlySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyr 57
           :::   |||:::      |||:::   :::      |||::::::::
   13231 CCAGTCATGAATCT......TGCAAACAGCATTACCCATTTAACAAATGG 13274
```

```
     58 Arg......AlaCysProCysLeuArgAsnLeuThr...CysIleTyrSe 71
        |||         |||   |||::::::|||:::::   |||   |||||
  13275 AGGCGCCAAATATGCAAATGCGTACAAAACATTTCCAATTGCAGATATTC 13324

     71 r...LysAsnGluLysTrpLeuSer 78
        |   ::::::|||   |||:::|||
  13325 CACAAGCACAGAATATTGGTTTTCT 13349

seq_name: gb_pr10:HSY10344

seq_documentation_block:
LOCUS       HSY10344     2734 bp    DNA             PRI       27-OCT-2000
DEFINITION  H.sapiens GalNAc-T2 gene, 3'UTR.
ACCESSION   Y10344
VERSION     Y10344.1  GI:2292904
KEYWORDS    GalNAc-T2 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2734)
  AUTHORS   Bennett,E.P., Weghuis,D.O., Merkx,G., van Kessel,A.G., Eiberg,H.
            and Clausen,H.
  TITLE     Genomic organization and chromosomal localization of three members
            of the UDP-N-acetylgalactosamine: polypeptide
            N-acetylgalactosaminyltransferase family
  JOURNAL   Glycobiology 8 (6), 547-555 (1998)
  MEDLINE   98256183
REFERENCE   2  (bases 1 to 2734)
  AUTHORS   Bennett,E.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-1997) E.P. Bennett, School Of Dentistry, Histo
            Lab, Noerre Alle 20, Copenhagen, 2200N, DENMARK
FEATURES             Location/Qualifiers
     source          1. .2734
                     /organism="Homo sapiens"
                     /isolate="P1 1732"
                     /db_xref="taxon:9606"
                     /clone_lib="P1"
     gene            1. .2734
                     /gene="GalNAc-T2"
     3'UTR           4. .2734
                     /gene="GalNAc-T2"
     polyA_signal    2633. .2638
                     /gene="GalNAc-T2"
                     /note="putative (box I)"
BASE COUNT      553 a    762 c    748 g    671 t
ORIGIN


alignment_scores:
           Quality:   87.00            Length:     78
             Ratio:  1.977              Gaps:       4
 Percent Similarity:  56.410   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179_COPY_24_121 x HSY10344/rev   ..

 Align seg 1/1  to reverse of: HSY10344  from: 1  to: 2734

      20 GlnAspCysGluThrGlyCysCys.........GlnArgAlaProAspAs 33
         ||||||||||||   |||   |||         ::.      |||   ::
    1207 CAGGACTGTGAGCCAGGCCTGTGCCCTGGCTGACGGGCTCATCCCTGACA 1158

      33 nCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnT 50
         :|||              :::|||||||||       |||:::
    1157 CTGCTG.............GCCATTAATTCTGAAGGCCCAGCGTGCAGGG 1121

      50 hrGlnValPhePheGlyGlnTyrArgAlaCys...ProCysLeuArgAsn 65
         ::|||:::   :::          ||||||     ||||||   :::
    1120 CTCAGATCACAATGAAGCCAATACTAGCGTGTGACAGCTGCCTGGCATCA 1071

      66 LeuThrCysIleTyrSerLysAsnGluLys......TrpLeuSerIleAl 80
         |||:::|||      :::::::::::::::      ||||||:::   ::
    1070 TTGGTCTGCCCTGTCGCCAGGGGGCAGCGGTATTGCTGGCTCACACAGAG 1021

      80 aTyrGlyArgCysGlnLysIleGlyArgGlnLys 91
         :||||||   |||      :::   |||:::|||
    1020 CTATGGGGTATGTCTTATATTCCTTAGAGAAAAG 987
```

```
OM of: US-09-215-435-179_COPY_24_121 to: N_Geneseq_0601:*   out_format : pfs

Date: Nov 15, 2001  4:28 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
 -DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000



score_list:
Sequence                                  Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614 +   547.00 1120.56  1.9e-54   542  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408 +   547.00 1118.23  2.6e-54   680  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465 +   500.00 1025.77  3.6e-49   421  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ41401 +   135.00  265.03  8.6e-07   656  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98928 +   135.00  264.88  8.8e-07   666  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08843 +    82.50  156.78  0.9215    624  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV07909 +    82.50  155.57  1.08      702  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75134 +    82.50  155.57  1.08      702  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56830 +    82.50  154.65  1.21      768  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08846 +    82.50  154.53  1.23      777  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08841 +    82.50  154.49  1.24      780  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08842 +    82.50  154.49  1.24      780  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75139 +    82.50  154.37  1.25      789  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56828 +    82.50  153.24  1.45      881  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA37042 +    82.50  140.12  7.81     3170
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF54250 +    82.50  140.12  7.81     3170
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75138 +    82.50  138.54  9.56     3696
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA44777 +    82.00  162.53  0.4410    322  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX22115 -    82.00  143.88  4.82     1984
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37622 -    79.50  149.08  2.47      721  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF74198 +    78.00  146.53  3.43      683  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ87844 +    77.50  138.72  9.35     1323
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA54529 -    76.00  135.03 15.00     1400
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35367 -    76.00  127.57 39.04     2898
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ52675 +    76.00  126.47 44.95     3226
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28470 -    76.00  124.30 59.41     3988
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28469 -    76.00  123.15 68.82     4460
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37621 -    75.50  143.46  5.09      556  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68221 +    75.00  131.87 22.48     1556
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV63754 +    75.00  121.42 85.95     4315
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX55560 +    75.00  120.52 96.36     4707
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC63721 +    75.00  120.52 96.36     4707
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF76945 +    75.00  120.52 96.36     4707
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF25249 +    75.00  120.52 96.36     4707
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC16991 +    74.00  142.34  5.87      458  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ68222 +    74.00  129.69 29.74     1573
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX30093 +    74.00  122.80 71.93     3079
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94993 +    74.00  122.80 71.93     3079
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94986 +    74.00  120.12 101.44    3999
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAD00679 +    73.50  140.59  7.35      491  !
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV07908 +    73.00  127.91  37.38     152
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ52674 +    73.00  117.98  133.63    402
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX30095 +    73.00  117.98  133.63    402
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX30094 +    73.00  117.89  135.02    406
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT67187 -    73.00  116.19  168.04    479

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97614

seq_documentation_block:
ID   AAX97614 standard; DNA; 542 BP.
XX
AC   AAX97614;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 78.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY35930.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 212; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 542 BP; 132 A; 146 C; 126 G; 138 T; 0 other;


alignment_scores:
           Quality:  547.00              Length:     98
             Ratio:  5.582                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAX97614   ..

 Align seg 1/1  to: AAX97614  from: 1  to: 542
```

```
      1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     85 TCACTTTCTCCAACAAAATATAACCTTTTGGAGCTCAAGGAGTCTTGCAT 134

     17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    135 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 184

     34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    185 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 234

     51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    235 CAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 284

     67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    285 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 334

     84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
        ||||||||||||||||||||||||||||||||||||||||||
    335 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 378
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA26408

seq_documentation_block:
ID AAA26408 standard; cDNA; 680 BP.
XX
AC AAA26408;
XX
DT 29-JUN-2000 (first entry)
XX
DE Human secreted protein gene 63 SEQ ID NO:73.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.
XX
OS Homo sapiens.
XX
PN WO200006698-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US17130.
XX
PR 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
DR WPI; 2000-195282/17.
DR P-PSDB; AAY91513.
XX
PT New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 416; 634pp; English.
XX
CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC corresponding secreted proteins are useful for preventing, treating or
CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC pathological conditions can be diagnosed by determining the amount of the
CC proteins in a sample or by determining the presence of mutations in the
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides, based on which tissues they are most highly expressed
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The proteins
CC or polynucleotides can also be used as food additives or preservatives.
CC The proteins are also useful for identifying their binding partners.
CC AAA26337 to AAA26345 and AAY91450 are sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 680 BP; 163 A; 185 C; 167 G; 156 T; 9 other;

alignment_scores:
Quality: 547.00 Length: 98
Ratio: 5.582 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-179_COPY_24_121 x AAA26408 ..

Align seg 1/1 to: AAA26408 from: 1 to: 680

```
      1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 251

     17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 301

     34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 351

     51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 CAGGTGTTCTTTGGCCAGTATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 401

     67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 451

     84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
        ||||||||||||||||||||||||||||||||||||||||||
    452 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 495
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX51465

seq_documentation_block:
ID AAX51465 standard; cDNA; 421 BP.
XX
AC AAX51465;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:44.
XX

```
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12687.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 177; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 421 BP; 96 A; 106 C; 110 G; 102 T; 7 other;


alignment_scores:
            Quality:  500.00              Length:     95
              Ratio:  5.435                 Gaps:      0
 Percent Similarity:  96.842   Percent Identity: 94.737

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAX51465   ..

 Align seg 1/1  to: AAX51465  from: 1  to: 421

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||:::
     131 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGKSKKGCAT 180

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
            ||||||||||||||||||||||||||||||||||||||||||||||||
     181 SGGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 230

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 280

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 330

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 380

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLys 95
         ||||||||||||||||||||||||||||||:::|||
     381 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAGRAAA 415

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ41401

seq_documentation_block:
ID   AAZ41401 standard; cDNA; 656 BP.
XX
AC   AAZ41401;
XX
DT   19-JAN-2000  (first entry)
XX
DE   Human normal pancreas tissue derived cDNA 16.
XX
KW   Human; pancreas; cancer; treatment; anticancer; cytostatic;
KW   gene therapy; EST; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   DE19818598-A1.
XX
PD   21-OCT-1999.
XX
PF   19-APR-1998;   98DE-1018598.
XX
PR   19-APR-1998;   98DE-1018598.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-592039/51.
XX
PT   New nucleic acid sequences expressed in normal pancreatic tissues, and
PT   derived polypeptides, for treatment of pancreatic cancer and
PT   identification of therapeutic agents  -
XX
PS   Claim 3; Page 56; 92pp; German.
XX
CC   This invention describes novel cDNA sequences (A) that are highly
CC   expressed in normal human pancreatic tissue and which have anticancer
CC   and cytostatic activity. (A) are used (i) for recombinant expression of
CC   polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to
CC   identify agents suitable for treatment of pancreatic cancer; (ii)
CC   directly for treating this form of cancer (including expression from
CC   gene therapy vectors) and (iii) for generation of specific antibodies.
CC   (A) are identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, so
CC   should reduce the number of failures associated with the fact that ESTs
CC   from different libraries may represent different parts of the same
CC   unknown gene, distorting the estimated frequency of occurrence in a
CC   particular tissue. AAZ41386-Z41423 represent human normal pancreatic
CC   tissue derived cDNA fragments which encode the protein fragments
CC   represented in AAY59893-Y59920.
XX
SQ   Sequence 656 BP; 165 A; 200 C; 162 G; 129 T; 0 other;


alignment_scores:
            Quality:  135.00              Length:     87
              Ratio:  2.328                 Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
```

```
US-09-215-435-179_COPY_24_121 x AAZ41401   ..

Align seg 1/1  to: AAZ41401  from: 1  to: 656

     10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
        ||||||      |||   |||:::::::::   :::|||:::::::::|| 
    170 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 219

     26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
        |||||||:::::              ::::::|||::::::   :::| 
    220 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 269

     43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
        |||||:::|||   |||:::::::::::::::::::|||   |||     
    270 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 319

     60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
        |||||||||   |||:::|||||||||         :::::::::|||  
    320 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 360

     76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
        :::             :::     :::|||   |||:::      |    
    361 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 410

     88 lyArgGlnLys 91
        |||||:::||| 
    411 GACGCTCCAAG 421

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98928

seq_documentation_block:
ID   AAC98928 standard; cDNA; 666 BP.
XX
AC   AAC98928;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:156.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54163.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 614; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 666 BP; 151 A; 217 C; 157 G; 133 T; 8 other;

alignment_scores:
            Quality:  135.00              Length:     87
              Ratio:  2.328                 Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAC98928   ..

 Align seg 1/1  to: AAC98928  from: 1  to: 666

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| 
     163 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 212

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||:::::              ::::::|||::::::   :::| 
     213 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 262

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::::|||   |||     
     263 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 312

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         :::::::::|||  
     313 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 353

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
         :::             :::     :::|||   |||:::      |    
     354 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 403

      88 lyArgGlnLys 91
         |||||:::||| 
     404 GACGCTCCAAG 414

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08843

seq_documentation_block:
ID   AAA08843 standard; DNA; 624 BP.
XX
AC   AAA08843;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human DKR-2 gene splice variant, DKR-2a.
XX
KW   DKR-2a; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
KW   growth factor; cytostatic; sonic hedgehog; tissue differentiation; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200018914-A2.
XX
PD   06-APR-2000.
XX
PF   17-SEP-1999;   99WO-US21647.
XX
PR   25-SEP-1998;   98US-0161241.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Bass MB,  Sullivan JK,  Theill LE,  Wang D;
XX
DR   WPI; 2000-293153/25.
DR   P-PSDB; AAY92074.
XX
PT   New nucleic acid molecule encoding a biologically active DKR
PT   polypeptide, useful in treatment of cancer, e.g. mammary tumors and
PT   stem cell tumors
XX
PS   Claim 1; Page 124; 143pp; English.
XX
CC   AAA08838-44 are novel mouse and human genes encoding DKR polypeptides.
CC   The human DKR-2a gene is a splice variant of DKR-2 and seems to be
CC   missing an exon.  Human DKR-2 has about 95 percent homology to murine
CC   DKR-2.  DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene
CC   identified in Xenopus and mouse, purportedly an antagonist of wnt-8
CC   signaling.  DKR-2, -3 and -4 are each related to DKR-1 by their cysteine
CC   pattern.  Dkk-1 is also involved in morphogenesis in the developing
CC   embryo, and therefore a growth factor, by inference DKR polypeptides
CC   are also growth factors.  The DKR polypeptides are useful for treating
CC   cancer, e.g. mammary tumors, stem cell tumors, or other cancers in which
CC   the wnt and/or sonic hedgehog (shh) signal transduction pathways are
CC   activated.  They can also be used to enhance tissue differentiation,
CC   such as bone formation and hematopoietic cell formation.
XX
SQ   Sequence 624 BP; 150 A; 161 C; 184 G; 129 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAA08843   ..

 Align seg 1/1  to: AAA08843  from: 1  to: 624

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::  |||:::|||:::::|||||      |||   |||
     385 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 426

      30 aProAspAsnCysGluSerHis.............CysAlaGluLysG 42
                 |||      |||              |||
     427 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 466

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||:::::::|||      |||         |||:::
     467 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 516

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   ::::::|||:::|||
     517 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 561

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                 :::   |||   :::   :::|||:::
     562 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 601

      84 ........CysGlnLysIle 87
                ||||||||||||
     602 TCCATGTGTGTCAGAAAATT 621

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV07909


seq_documentation_block:
ID   AAV07909 standard; cDNA; 702 BP.
XX
AC   AAV07909;
XX
DT   18-JAN-1999  (first entry)
XX
DE   Human cysteine-rich secreted protein CRSP-4 cDNA.
XX
KW   CRSP-4; cysteine-rich secreted protein 4; tumour; cancer; leukaemia;
KW   tissue repair; wound healing; infection; Parkinson's disease;
KW   Alzheimer's disease; Huntington's chorea; multiple sclerosis;
KW   amyotrophic lateral sclerosis; pontine myelinolysis;
KW   human immunodeficiency associated myelopathy; bulbar palsy;
KW   spinal muscular atrophy; primary lateral sclerosis; poliomyelitis;
KW   Fazio-Londe syndrome; Charcot-Marie-Tooth disease; therapy;
KW   diagnosis; drug screening; human; signal transduction;
KW   cell differentiation; cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..540
FT                   /*tag=  a
FT                   /note= "claimed fragment"
XX
PN   WO9846755-A1.
XX
PD   22-OCT-1998.
XX
PF   16-APR-1998;   98WO-US07894.
XX
PR   20-JAN-1998;   98US-0009802.
PR   16-APR-1997;   97US-0843704.
PR   17-APR-1997;   97US-0842898.
PR   15-JAN-1998;   98US-0071589.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   McCarthy SA;
XX
DR   WPI; 1998-568730/48.
DR   P-PSDB; AAW73019.
XX
PT   New isolated cysteine-rich secreted proteins - used to develop
PT   products for treating, e.g. hyperproliferative disorders, cancers,
PT   wounds, infectious lesions, degenerative lesions or demyelating
PT   diseases
XX
PS   Claim 1; Page 104-105; 142pp; English.
XX
CC   This partial cDNA clone, the coding region for which is also
CC   claimed, codes for novel human cysteine-rich secreted protein 4
CC   (CRSP-4, see AAW73019).  A partial sequence was derived from dbEST
CC   clone AAW55979  using human CRSP-1 as probe.  This clone was
CC   subsequently obtained from the IMAGE collection and sequenced fully
CC   to define the CRSP-4 sequence.  CRSP-4 mRNA is expressed in several
CC   adult human tissues, with highest expression in hear, brain,
CC   placenta, lung and skeletal muscle.  CRSP-1, CRSP-2 CRSP-3 and
CC   CRSP-4 nucleic acid molecules (see AAV07906-09) and polypeptides (see
CC   AAW73016-19) are claimed.  The CRSPs have at least 1 of the following
CC   activities: (i) modulation of cellular signal transduction, either
CC   in vitro or in vivo (e.g. antagonism of the activity of members of
CC   the wnt family of secreted proteins or suppression of wnt-dependent
CC   signal transduction); (ii) regulation of communication between cells
CC   (e.g. regulation of wnt-dependent cell-cell interactions); (iii)
CC   regulation of expression of genes whose expression is modulated by
CC   binding of CRSP (e.g. CRSP-1) to a receptor; (iv) regulation of gene
CC   transcription in a cell involved in development or differentiation,
CC   either in vitro or in vivo (e.g. induction of cell differentiation);
CC   (v) regulation of cellular proliferation, either in vitro or in vivo
CC   (e.g. induction of cellular proliferation or inhibition of
```

```
CC   proliferation as in the case of suppression of tumurogenesis (e.g.
CC   suppression of glioblastoma formation)); (vi) formation and
CC   maintenance of ordered spatial arrangements of differentiated
CC   tissues in vertebrates, both adult and embryonic (e.g. induction of
CC   head formation during vertebrate development or maintenance of
CC   haematopoietic progenitor cells); (vii) modulation of cell death,
CC   such as stimulation of cell survival; (ix) regulating cell
CC   migration, and/or (xii) immune modulation.  Products of the
CC   invention can be used for e.g. treating hyperproliferative diseases
CC   such as neoplastic and hyperplastic disease, e.g. various cancers
CC   and leukaemias, and fibroproliferative disorders, malignancy,
CC   premalignant and benign conditions, to stimulate tissue repair or
CC   wound healing or trea spinal cord injuries, brain injuries, lesions
CC   associated with surgery, ischemic lesions, malignant lesions,
CC   infectious lesions, degenerative lesions (Parkinson's disease,
CC   Alzheimer's disease, Huntington's chorea, amyotrophic lateral
CC   sclerosis), demyelating diseases (multiple sclerosis, human
CC   immunodeficiency associated myelopathy, pontine myelinolysis),
CC   motor neuron injuries, progressive spinal muscular atrophy,
CC   progressive bulbar palsy, primary lateral sclerosis, infantile and
CC   juvenile muscular atrophy, progressive bulbar paralysis of childhood
CC   (Fazio-Londe syndrome), poliomyelitis, and hereditary motorsensory
CC   neuropathy (Charcot-Marie-Tooth disease).  The products can also be
CC   used for detection, diagnosis and drug screening.
XX
SQ   Sequence 702 BP; 221 A; 158 C; 171 G; 152 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAV07909   ..

 Align seg 1/1  to: AAV07909  from: 1  to: 702

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::||||||      |||   |||
     301 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 342

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||               |||
     343 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 382

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||::::::|||      |||         |||:::
     383 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 432

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   :::::::|||:::|||
     433 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 477

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
               :::   |||   :::   :::|||:::
     478 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 517

      84 ........CysGlnLysIle 87
                |||||||||||
     518 TCCATGTGTGTCAGAAAATT 537

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75134

seq_documentation_block:
ID   AAA75134 standard; cDNA; 702 BP.
XX
AC   AAA75134;
XX
DT   15-JAN-2001  (first entry)
XX
DE   cDNA encoding a partial human Dickkopf (Dkk)-2 protein.
XX
KW   Human; Dickkopf-2 protein; Dkk-2 protein; Soggy protein; optic disorder;
KW   cysteine-rich secreted protein; glaucoma; conjunctivitis; brain disorder;
KW   Alzheimer's disease; epilepsy; amnesia; inflammation; pulmonary disorder;
KW   skeletal muscle disorder; Goodpasture's syndrome;
KW   cardiovascular disorder; hyperproliferative disorder; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..540
FT                   /*tag= a
FT                   /product= " Dickkopf (Dkk)-2 protein"
XX
PN   WO200052047-A2.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-US05452.
XX
PR   05-MAR-1999;   99US-0263022.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   McCarthy S;
XX
DR   WPI; 2000-579276/54.
DR   P-PSDB; AAB08877.
XX
PT   Human Dickkopf (hDkk) and Soggy nucleic acids and proteins, useful as
PT   modulating agents in regulating cellular processes and particularly for
PT   treating disorders characterized by aberrant expression or activity of
PT   Dkk, e.g. Alzheimer's, -
XX
PS   Example 6; Page 81-82; 208pp; English.
XX
CC   The present sequence encodes a partial Dickkopf (Dkk)-2 protein. The
CC   specification also describes Soggy (Dkk-related) sequences. Dkk is a
CC   cysteine-rich secreted protein. The Dkk nucleic acids and proteins are
CC   useful as modulating agents in regulating cellular processes. They are
CC   particularly useful in treating subjects having a disorder characterized
CC   by aberrant expression or activity of Dkk such as optic disorders
CC   (glaucoma, conjunctivitis), brain disorders (Alzheimer's disease,
CC   epilepsy, amnesia), inflammation, skeletal muscle disorders, pulmonary
CC   disorders (Goodpasture's syndrome), cardiovascular disorders, and
CC   hyperproliferative disorders (cancer). The Dkk proteins and nucleic
CC   acids may also be used for research purposes, such as for chromosome
CC   mapping, tissue typing and in screening assays to identify modulators.
XX
SQ   Sequence 702 BP; 221 A; 158 C; 171 G; 152 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAA75134   ..

 Align seg 1/1  to: AAA75134  from: 1  to: 702

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::||||||      |||   |||
     301 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 342

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||               |||
     343 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 382

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln......'56
            :::|||::::::|||      |||         |||:::
     383 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 432
```

```
      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   :::::::|||:::|||
     433 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 477

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                  :::   |||   :::   :::|||:::
     478 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 517

      84 ........CysGlnLysIle 87
                 ||||||||||||
     518 TCCATGTGTGTCAGAAAATT 537

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56830

seq_documentation_block:
ID   AAX56830 standard; cDNA; 768 BP.
XX
AC   AAX56830;
XX
DT   14-JUL-1999  (first entry)
XX
DE   Human phdkk-2 cDNA.
XX
KW   Signal pathway; wnt; inhibitor; secreted glycoprotein; receptor;
KW   therapy; diagnosis; treatment; cancer; breast; carcinoma; melanoma;
KW   colon; cell proliferation; differentiation; ss.
XX
OS   Homo sapiens.
XX
PN   WO9922000-A1.
XX
PD   06-MAY-1999.
XX
PF   27-OCT-1998;   98WO-DE03155.
XX
PR   27-OCT-1997;   97DE-1047418.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Glinka A,  Niehrs C;
XX
DR   WPI; 1999-303017/25.
XX
PT   Protein that inhibits wnt signalling
XX
PS   Claim 4a; Fig 2; 39pp; German.
XX
CC   This invention describes an inhibitory protein of the wnt signalling
CC   pathway (which comprises secreted glycoproteins and their associated
CC   receptors). The products of the invention are useful as therapeutic
CC   and diagnostic agents, e.g. to treat cancer (e.g. carcinoma of breast
CC   and colon, or melanoma) and for studying processes associated with wnt
CC   signalling (cell proliferation and differentiation). Primers derived
CC   from the nucleic acid of the invention are used to detect the expression
CC   of the gene. AAX56827-X26833 are sequences used in the method of the
CC   invention.
XX
SQ   Sequence 768 BP; 239 A; 172 C; 186 G; 171 T; 0 other;


alignment_scores:
            Quality:   82.50            Length:     90
              Ratio:  2.062              Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAX56830   ..

 Align seg 1/1  to: AAX56830  from: 1  to: 768

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::||||||      |||   |||
     339 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 380

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||                |||
     381 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 420

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||::::::|||      |||        |||:::
     421 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 470

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   :::::::|||:::|||
     471 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 515

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                  :::   |||   :::   :::|||:::
     516 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 555

      84 ........CysGlnLysIle 87
                 ||||||||||||
     556 TCCATGTGTGTCAGAAAATT 575

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08846

seq_documentation_block:
ID   AAA08846 standard; DNA; 777 BP.
XX
AC   AAA08846;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human DKR-2 DNA optimized for expression in E. coli.
XX
KW   DKR-2; recombinant; bacterial expression; dkk-1; dickkopf-1; antagonist;
KW   wnt-8 signaling; morphogenesis; growth factor; cytostatic;
KW   sonic hedgehog; tissue differentiation; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200018914-A2.
XX
PD   06-APR-2000.
XX
PF   17-SEP-1999;   99WO-US21647.
XX
PR   25-SEP-1998;   98US-0161241.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Bass MB,  Sullivan JK,  Theill LE,  Wang D;
XX
DR   WPI; 2000-293153/25.
XX
PT   New nucleic acid molecule encoding a biologically active DKR
PT   polypeptide, useful in treatment of cancer, e.g. mammary tumors and
PT   stem cell tumors
XX
PS   Claim 1; Page 142-143; 143pp; English.
XX
CC   AAA08845-48 are recombinant human DKR constructs with codons optimized
CC   for expression in Escherichia coli.  Recombinant DKR-2 was engineered to
CC   encode a 235 amino acid polypeptide.
CC   DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene
CC   identified in Xenopus and mouse, purportedly an antagonist of wnt-8
CC   signaling.  DKR-2, -3 and -4 are each related to DKR-1 by their cysteine
CC   pattern.  Dkk-1 is also involved in morphogenesis in the developing
CC   embryo, and therefore a growth factor, by inference DKR polypeptides
CC   are also growth factors.  The DKR polypeptides are useful for treating
CC   cancer, e.g. mammary tumors, stem cell tumors, or other cancers in which
CC   the wnt and/or sonic hedgehog (shh) signal transduction pathways are
CC   activated.  They can also be used to enhance tissue differentiation,
CC   such as bone formation and hematopoietic cell formation.
XX
```

```
SQ   Sequence 777 BP; 167 A; 245 C; 186 G; 179 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAA08846   ..

 Align seg 1/1  to: AAA08846  from: 1  to: 777

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::|||||      |||   |||
     541 GACCCGTGCCTGCGTTCCTCCGACTGCATCGAAGGTTTCTGC........ 582

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||                |||
     583 ..........TGCGCTCGTCACTTCTGGACCAAAATCTGCAAACCGGTTC 622

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||::::::|||      |||         |||:::
     623 TGCACCAGGGTGAAGTTTGCACCAAACAGCGTAAAAAAGGTTCCCACGGT 672

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   ::::::|||:::|||
     673 CTGGAAATCTTCCAGCGTTGCGACTGCGCTAAAGGTCTGTCCTGC..... 717

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                 :::   |||   :::   :::|||:::
     718 ..........AAAGTTTGGAAAGACGCTACCTACTCCTCCAAAGCTCGTC 757

      84 ........CysGlnLysIle 87
                 ||||||||||||
     758 TGCACGTTTGCCAGAAAATC 777

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08841

seq_documentation_block:
ID   AAA08841 standard; DNA; 780 BP.
XX
AC   AAA08841;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Murine DKR-2 DNA.
XX
KW   DKR-2; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
KW   growth factor; cytostatic; sonic hedgehog; tissue differentiation; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..99
FT                   /*tag=  a
FT   mat_peptide     100..778
FT                   /*tag=  b
XX
PN   WO200018914-A2.
XX
PD   06-APR-2000.
XX
PF   17-SEP-1999;   99WO-US21647.
XX
PR   25-SEP-1998;   98US-0161241.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Bass MB,  Sullivan JK,  Theill LE,  Wang D;
XX
DR   WPI; 2000-293153/25.
DR   P-PSDB; AAY92072.
XX
PT   New nucleic acid molecule encoding a biologically active DKR
PT   polypeptide, useful in treatment of cancer, e.g. mammary tumors and
PT   stem cell tumors
XX
PS   Claim 1; Page 122; 143pp; English.
XX
CC   AAA08838-44 are novel mouse and human genes encoding DKR polypeptides.
CC   The murine DKR-2 gene has about 38 percent homology to murine dkk-1
CC   (dickkopf-1), a novel gene identified in Xenopus and mouse, purportedly
CC   an antagonist of wnt-8 signaling.
CC   Dkk-1 is also involved in morphogenesis in the developing embryo, and
CC   therefore a growth factor, by inference DKR polypeptides are also
CC   growth factors.  The DKR polypeptides are useful for treating cancer,
CC   e.g. mammary tumors, stem cell tumors, or other cancers in which the wnt
CC   and/or sonic hedgehog (shh) signal transduction pathways are activated.
CC   They can also be used to enhance tissue differentiation, such as bone
CC   formation and hematopoietic cell formation.
XX
SQ   Sequence 780 BP; 205 A; 212 C; 209 G; 154 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAA08841   ..

 Align seg 1/1  to: AAA08841  from: 1  to: 780

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::|||||      |||   |||
     541 GACCCATGCCTACGGTCATCAGACTGCATTGATGGGTTTTGT........ 582

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||                |||
     583 ..........TGTGCTCGCCACTTCTGGACCAAAATCTGCAAACCAGTGC 622

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||::::::|||      |||         |||:::
     623 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCGCACGGG 672

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   ::::::|||:::|||
     673 CTGGAGATTTTCCAGAGGTGTGACTGTGCAAAGGGCCTGTCCTGC..... 717

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                 :::   |||   :::   :::|||:::
     718 ..........AAAGTGTGGAAAGATGCCACCTACTCTTCCAAAGCCAGAC 757

      84 ........CysGlnLysIle 87
                 ||||||||||||
     758 TCCATGTATGCCAGAAGATC 777

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA08842

seq_documentation_block:
ID   AAA08842 standard; DNA; 780 BP.
XX
AC   AAA08842;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human DKR-2 DNA.
XX
KW   DKR-2; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
KW   growth factor; cytostatic; sonic hedgehog; tissue differentiation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
```

```
FT   sig_peptide     1..99
FT                   /*tag=  a
FT   mat_peptide     100..778
FT                   /*tag=  b
XX
PN   WO200018914-A2.
XX
PD   06-APR-2000.
XX
PF   17-SEP-1999;   99WO-US21647.
XX
PR   25-SEP-1998;   98US-0161241.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Bass MB,  Sullivan JK,  Theill LE,  Wang D;
XX
DR   WPI; 2000-293153/25.
DR   P-PSDB; AAY92073.
XX
PT   New nucleic acid molecule encoding a biologically active DKR
PT   polypeptide, useful in treatment of cancer, e.g. mammary tumors and
PT   stem cell tumors
XX
PS   Claim 1; Page 122-123; 143pp; English.
XX
CC   AAA08838-44 are novel mouse and human genes encoding DKR polypeptides.
CC   The human DKR-2 gene has about 95 percent homology to murine DKR-2.
CC   DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene identified
CC   in Xenopus and mouse, purportedly an antagonist of wnt-8 signaling.
CC   DKR-2, -3 and -4 are each related to DKR-1 by their cysteine pattern.
CC   Dkk-1 is also involved in morphogenesis in the developing embryo, and
CC   therefore a growth factor, by inference DKR polypeptides are also
CC   growth factors.  The DKR polypeptides are useful for treating cancer,
CC   e.g. mammary tumors, stem cell tumors, or other cancers in which the wnt
CC   and/or sonic hedgehog (shh) signal transduction pathways are activated.
CC   They can also be used to enhance tissue differentiation, such as bone
CC   formation and hematopoietic cell formation.
XX
SQ   Sequence 780 BP; 200 A; 203 C; 215 G; 162 T; 0 other;


alignment_scores:
            Quality:   82.50            Length:     90
              Ratio:  2.062              Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAA08842   ..

 Align seg 1/1  to: AAA08842  from: 1  to: 780

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::||||||      |||   |||
     541 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 582

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||               |||
     583 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 622

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||:::::|||      |||         |||:::
     623 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 672

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   :::::::|||:::|||
     673 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 717

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                 :::   |||   :::   :::|||:::
     718 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 757

      84 ........CysGlnLysIle 87
```

```
                 ||||||||||||
     758 TCCATGTGTGTCAGAAAATT 777

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA75139

seq_documentation_block:
ID   AAA75139 standard; cDNA; 789 BP.
XX
AC   AAA75139;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Open reading frame encoding a human Dickkopf (Dkk)-2 protein.
XX
KW   Human; Dickkopf-2 protein; Dkk-2 protein; Soggy protein; optic disorder;
KW   cysteine-rich secreted protein; glaucoma; conjunctivitis; brain disorder;
KW   Alzheimer's disease; epilepsy; amnesia; inflammation; pulmonary disorder;
KW   skeletal muscle disorder; Goodpasture's syndrome;
KW   cardiovascular disorder; hyperproliferative disorder; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..789
FT                   /*tag=  a
FT                   /product= " Dickkopf (Dkk)-2 protein"
FT   sig_peptide     1..102
FT                   /*tag=  b
XX
PN   WO200052047-A2.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-US05452.
XX
PR   05-MAR-1999;   99US-0263022.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   McCarthy S;
XX
DR   WPI; 2000-579276/54.
DR   P-PSDB; AAB08880.
XX
PT   Human Dickkopf (hDkk) and Soggy nucleic acids and proteins, useful as
PT   modulating agents in regulating cellular processes and particularly for
PT   treating disorders characterized by aberrant expression or activity of
PT   Dkk, e.g. Alzheimer's, -
XX
PS   Claim 2; Page 97-98; 208pp; English.
XX
CC   The present sequence encodes a human Dickkopf (Dkk)-2 protein. The
CC   specification also describes Soggy (Dkk-related) sequences. Dkk is a
CC   cysteine-rich secreted protein. The Dkk nucleic acids and proteins are
CC   useful as modulating agents in regulating cellular processes. They are
CC   particularly useful in treating subjects having a disorder characterized
CC   by aberrant expression or activity of Dkk such as optic disorders
CC   (glaucoma, conjunctivitis), brain disorders (Alzheimer's disease,
CC   epilepsy, amnesia), inflammation, skeletal muscle disorders, pulmonary
CC   disorders (Goodpasture's syndrome), cardiovascular disorders, and
CC   hyperproliferative disorders (cancer). The Dkk proteins and nucleic
CC   acids may also be used for research purposes, such as for chromosome
CC   mapping, tissue typing and in screening assays to identify modulators.
XX
SQ   Sequence 789 BP; 200 A; 208 C; 219 G; 162 T; 0 other;


alignment_scores:
            Quality:   82.50            Length:     90
              Ratio:  2.062              Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
```

```
US-09-215-435-179_COPY_24_121 x AAA75139   ..

Align seg 1/1  to: AAA75139  from: 1  to: 789

     14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
        :::   |||:::|||::::::|||||||      |||   |||
    553 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC........ 594

     30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                |||      |||                |||
    595 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 634

     42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
           :::|||::::::|||      |||         |||:::
    635 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 684

     57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                 ::::::   |||   |||   ::::::|||:::|||
    685 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 729

     70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                :::   |||   :::   :::|||:::
    730 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 769

     84 ........CysGlnLysIle 87
                ||||||||||||
    770 TCCATGTGTGTCAGAAAATT 789

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX56828

seq_documentation_block:
ID   AAX56828 standard; cDNA; 881 BP.
XX
AC   AAX56828;
XX
DT   14-JUL-1999  (first entry)
XX
DE   Mouse pmdkk-2 cDNA.
XX
KW   Signal pathway; wnt; inhibitor; secreted glycoprotein; receptor;
KW   therapy; diagnosis; treatment; cancer; breast; carcinoma; melanoma;
KW   colon; cell proliferation; differentiation; ss.
XX
OS   Mus sp.
XX
PN   WO9922000-A1.
XX
PD   06-MAY-1999.
XX
PF   27-OCT-1998;   98WO-DE03155.
XX
PR   27-OCT-1997;   97DE-1047418.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Glinka A,  Niehrs C;
XX
DR   WPI; 1999-303017/25.
XX
PT   Protein that inhibits wnt signalling
XX
PS   Claim 4a; Fig 2; 39pp; German.
XX
CC   This invention describes an inhibitory protein of the wnt signalling
CC   pathway (which comprises secreted glycoproteins and their associated
CC   receptors). The products of the invention are useful as therapeutic
CC   and diagnostic agents, e.g. to treat cancer (e.g. carcinoma of breast
CC   and colon, or melanoma) and for studying processes associated with wnt
CC   signalling (cell proliferation and differentiation). Primers derived
CC   from the nucleic acid of the invention are used to detect the expression
CC   of the gene. AAX56827-X26833 are sequences used in the method of the
CC   invention.
XX
```

```
SQ   Sequence 881 BP; 269 A; 197 C; 228 G; 187 T; 0 other;


alignment_scores:
            Quality:   82.50              Length:     90
              Ratio:  2.062                Gaps:      5
 Percent Similarity:  44.444   Percent Identity: 25.556

alignment_block:
 US-09-215-435-179_COPY_24_121 x AAX56828   ..

 Align seg 1/1  to: AAX56828  from: 1  to: 881

      14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
         :::   |||:::|||::::::|||||||      |||   |||
     378 GACCCATGCCTACGGTCATCAGACTGCATTGATGGGTTTTGT........ 419

      30 aProAspAsnCysGluSerHis..............CysAlaGluLysG 42
                 |||      |||                |||
     420 ..........TGTGCTCGCCACTTCTGGACCAAAATCTGCAAACCAGTGC 459

      42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln...... 56
            :::|||::::::|||      |||         |||:::
     460 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCGCACGGG 509

      57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
                  ::::::   |||   |||   ::::::|||:::|||
     510 CTGGAGATTTTCCAGAGGTGTGACTGTGCAAAGGGCCTGTCCTGC..... 554

      70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
                 :::   |||   :::   :::|||:::
     555 ..........AAAGTGTGGAAAGATGCCACCTACTCTTCCAAAGCCAGAC 594

      84 ........CysGlnLysIle 87
                 ||||||||||||
     595 TCCATGTATGCCAGAAGATC 614

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA37042

seq_documentation_block:
ID   AAA37042 standard; cDNA; 3170 BP.
XX
AC   AAA37042;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1316 (UNQ682) cDNA sequence SEQ ID NO:69.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
```

PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR P-PSDB; AAY99360.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 2; Fig 41; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 3170 BP; 1014 A; 613 C; 629 G; 914 T; 0 other;

alignment_scores:
Quality: 82.50 Length: 90
Ratio: 2.062 Gaps: 5
Percent Similarity: 44.444 Percent Identity: 25.556

alignment_block:
US-09-215-435-179_COPY_24_121 x AAA37042 ..

Align seg 1/1 to: AAA37042 from: 1 to: 3170

```
 14 GluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAl 30
    :::   |||:::|||:::::::||||||      |||   |||
751 GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGC,..... 792

 30 aProAspAsnCysGluSerHis.............CysAlaGluLysG 42
             |||      |||               |||
793 ..........TGTGCTCGTCATTTCTGGACCAAAATCTGCAAACCAGTGC 832

 42 lySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGln..... 56
       :::|||:::::::|||      |||         |||:::
833 TCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAGGGTTCTCATGGG 882

 57 .........TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTy 70
            ::::::   |||   |||   ::::::|||:::|||
883 CTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC..... 927

 70 rSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArg.......... 83
             :::   |||   :::   :::|||:::
928 ..........AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGAC 967

 84 ........CysGlnLysIle 87
           ||||||||||||
968 TCCATGTGTGTCAGAAAATT 987
```

```
OM of: US-09-215-435-179_COPY_24_121 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  4:11 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2 +   86.50   79.17  9.2e+03   4403765
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-11 +   76.00  133.39  8.84      3226  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-11 +   76.00  133.39  8.84      3226  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-11 +   76.00  133.39  8.84      3226
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-3 +   75.00  139.01  4.30      1556  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-3 +   75.00  139.01  4.30      1556  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-882-046-3 +   75.00  128.05  17.55     4315  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-181-706-1 +   75.00  127.11  19.78     4707  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-458-791-1 +   75.00  127.11  19.78     4707  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-459-066-1 +   75.00  127.11  19.78     4707  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-4 +   74.00  136.67  5.81      1573  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-4 +   74.00  136.67  5.81      1573  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-9 +   73.00  124.34  28.23     4029  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-9 +   73.00  124.34  28.23     4029  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-9 +   73.00  124.34  28.23     4029
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-085-199B-3 -   73.00  122.47  35.89     4796  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-551-437-4 +   70.50  127.84  18.02     1734  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-004-225-4 +   70.50  127.84  18.02     1734  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-084-346-4 +   70.50  127.84  18.02     1734  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-104-704-4 +   70.50  127.84  18.02     1734  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-214-278-4 +   70.50  118.99  56.11     3955  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-400-159-7 +   70.50  117.69  66.29     4464  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-611-729A-7 +   70.50  117.64  66.68     4483  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-924-063A-4 +   70.00  142.60  2.72      396  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-085-199B-1 -   69.00  128.79  15.95     1164  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-188-930-255 +   69.00  126.33  21.88     1464  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-188-930-73 +   69.00  125.16  25.43     1633  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-13 +   68.00  121.79  39.15     1816  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-13 +   68.00  121.79  39.15     1816  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-13 +   68.00  121.79  39.15     1816
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-083-590A-21 +   67.50  102.66  455.55    9723  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-532-384-21 +   67.50  102.66  455.55    9723  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-226-264-27 +   67.00  130.06  13.56      684  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-454-455-3 +   67.00  114.28  102.61    2972  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-454-455-5 +   67.00  111.67  143.39    3789  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-537-361E-7 -   66.50  115.56  87.02     2378  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-525-940-22 +   66.00  114.37  101.46    2397  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-976-838-22 +   66.00  114.37  101.46    2397  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-525-940-20 +   66.00  113.32  116.07    2643  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-976-838-20 +   66.00  113.32  116.07    2643  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-525-940-19 +   66.00  112.91  122.29    2745
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-976-838-19 +   66.00  112.91  122.29    2745
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-525-940-17 +   66.00  112.83  123.58    2766
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-976-838-17 +   66.00  112.83  123.58    2766
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-976-838-28 -   66.00  112.83  123.58    2766

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
           Quality:   86.50            Length:     73
             Ratio:  2.403              Gaps:      5
 Percent Similarity:  49.315   Percent Identity: 30.137

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-103-840A-2   ..

 Align seg 1/1  to: US-09-103-840A-2  from: 1  to: 4403765

      16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
         |||   ||||||:::   |||:::      ||||||   |||:::|||
  3963587TGTTGCCGAAACCGATGTTGCCAATGCCGTTGTTGC...CGAAGCCCC.. 3963631

      32 pAsnCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysG 49
          :::|||      :::|||:::::::::
  3963632.AGTTGCTGCTGCCGGTGTTCCCGGCGC..................... 3963658

      49 lnThrGlnValPhePheGlyGlnTyrArgAlaCysProCys..... 62
                                 |||:::|||   |||
  3963659.......................CGATGTTGCCGCTGCCGGCGTTTC 3963682

      63 ...LeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLeuSe 78
            |||:::      |||   ||||||:::::::::::::|||   ||
  3963683CGGTGCCGACGTTGTTGTTGCCGGTATTCCCGAACCCGACGTTGG...AG 3963729

      78 rIleAlaTyrGlyArgCys 84
         |      |||   ||||||
  3963730TTGCCGGTATTGCCGCTGC 3963748

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-11

seq_documentation_block:
; Sequence 11, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
```

```
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  136..2543
US-07-862-021B-11


alignment_scores:
            Quality:   76.00              Length:    101
              Ratio:  1.407                Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-07-862-021B-11   ..

 Align seg 1/1  to: US-07-862-021B-11  from: 1  to: 3226

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||||||   |||:::::::|||:::   :::   ||
    1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

      27 sGlnArgAlaPro.............AspAsnCysGluSerHisCysAlaG 40
         |     :::                 |||:::|||:::::::   |||:::
    1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

      40 lu................................LysGlySerGluGly 45
                                           ::::::|||
    1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::       ::::::            |||
    1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
            :::::::::   |||:::   |||      :::::::|||
    1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::  ::::::      ||||||:::::
    1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

      94 Lys 94
         |||
    2044 AAA 2046

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-11

seq_documentation_block:
; Sequence 11, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  136..2543
US-08-313-288B-11


alignment_scores:
            Quality:   76.00              Length:    101
              Ratio:  1.407                Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-313-288B-11   ..

 Align seg 1/1  to: US-08-313-288B-11  from: 1  to: 3226

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||||||   |||:::::::|||:::   :::   ||
    1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

      27 sGlnArgAlaPro.............AspAsnCysGluSerHisCysAlaG 40
         |     :::                 |||:::|||:::::::   |||:::
    1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

      40 lu................................LysGlySerGluGly 45
                                           ::::::|||
    1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896
```

```
      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::             |||
    1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
            :::::::::   |||:::   |||       ::::::|||
    1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::   ::::::      ||||||::::::
    1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

      94 Lys 94
         |||
    2044 AAA 2046

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-11

seq_documentation_block:
; Sequence 11, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3226 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  136..2543
PCT-US93-03164-11


alignment_scores:
            Quality:   76.00              Length:    101
              Ratio:  1.407                Gaps:      4
 Percent Similarity:  53.465   Percent Identity: 21.782

alignment_block:
 US-09-215-435-179_COPY_24_121 x PCT-US93-03164-11    ..
```

```
 Align seg 1/1  to: PCT-US93-03164-11  from: 1  to: 3226

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::||||||   |||:::::::|||:::   :::   ||
    1747 GAAGAAACTGAGAAATGTATTGTAAATGAGGAATGCTCCCCTAGCAGCTG 1796

      27 sGlnArgAlaPro.............AspAsnCysGluSerHisCysAlaG 40
         |       :::                |||:::|||::::::   |||:::
    1797 CCTTGTCACCGAATGGGGAGAGTGGGATGAATGCAGTGCTAGCTGTGGCA 1846

      40 lu.....................................LysGlySerGluGly 45
                                                  ::::::|||
    1847 CAGGAATGAAAAGGCGACACAGAATGATCAAGATGACTCCTGCTGATGGA 1896

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::             |||
    1897 TCTATGTGCAAGGCAGAAACTACAGAGGCAGAGAAATGCATGATGCCCGA 1946

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
            :::::::::   |||:::   |||       ::::::|||
    1947 ATGCCATACTATTCCCTGCCTTCTATCCCCATGGTCTGAATGGAGCGACT 1996

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::   ::::::      ||||||::::::
    1997 GCAGCGTGACATGTGGGAAGGGAATGCGAACC...CGGCAAAGGATGCTG 2043

      94 Lys 94
         |||
    2044 AAA 2046

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-3

seq_documentation_block:
; Sequence 3, Application US/08597545
; Patent No. 5580738
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,545
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    ORIGINAL SOURCE:
;      STRAIN:  Human Dlk
US-08-597-545-3


alignment_scores:
            Quality:   75.00              Length:      61
              Ratio:  2.143                Gaps:       5
 Percent Similarity:  57.377   Percent Identity: 34.426

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-597-545-3   ..

 Align seg 1/1  to: US-08-597-545-3  from: 1  to: 1556

      12 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 28
         |||   :::::::|||:::   :::      |||      |||   |||
     330 CTTTGTGACCAGTGCGTGACCTCTCCCGGCTGCCTTCACGGACTCTGTGG 379

      28 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 45
          :::   |||   :::|||         |||::::::   |||   :::|
     380 AGAA...CCCGGGCAGTGC......ATTTGCACCGAC...GGCTGGGACG 417

      45 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 60
         ||:::||||||:::   :::|||              |||||||||
     418 GGGAGCTCTGTGATAGAGATGTT..............CGGGCCTGCTCC 452

      61 ......ProCysLeuArgAsnLeuThrCysIle 69
               ||||||   :::|||   ||||||:::
     453 TCGGCCCCCTGTGCCAACAACGGGACCTGCGTG 485

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-3

seq_documentation_block:
; Sequence 3, Application US/08457135
; Patent No. 5644031
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,135
;      FILING DATE:  01-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/304/NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
```

```
;      STRAIN:  Human Dlk
US-08-457-135-3


alignment_scores:
            Quality:   75.00              Length:      61
              Ratio:  2.143                Gaps:       5
 Percent Similarity:  57.377   Percent Identity: 34.426

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-457-135-3   ..

 Align seg 1/1  to: US-08-457-135-3  from: 1  to: 1556

      12 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 28
         |||   :::::::|||:::   :::      |||      |||   |||
     330 CTTTGTGACCAGTGCGTGACCTCTCCCGGCTGCCTTCACGGACTCTGTGG 379

      28 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 45
          :::   |||   :::|||         |||::::::   |||   :::|
     380 AGAA...CCCGGGCAGTGC......ATTTGCACCGAC...GGCTGGGACG 417

      45 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 60
         ||:::||||||:::   :::|||              |||||||||
     418 GGGAGCTCTGTGATAGAGATGTT..............CGGGCCTGCTCC 452

      61 ......ProCysLeuArgAsnLeuThrCysIle 69
               ||||||   :::|||   ||||||:::
     453 TCGGCCCCCTGTGCCAACAACGGGACCTGCGTG 485

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-882-046-3

seq_documentation_block:
; Sequence 3, Application US/08882046
; Patent No. 6136952
;  GENERAL INFORMATION:
;    APPLICANT:  Li, Linheng
;    APPLICANT:  Hood, Leroy
;    APPLICANT:  Krantz, Ian D.
;    APPLICANT:  Spinner, Nancy B.
;    TITLE OF INVENTION:  Human Jagged Polypeptide, Encoding
;    TITLE OF INVENTION:  Nucleic Acids and Methods of Use
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/882,046
;      FILING DATE:  25-JUN-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UW 2637
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4315 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  16..3460
US-08-882-046-3


alignment_scores:
            Quality:   75.00              Length:     99
              Ratio:  1.531                Gaps:      4
 Percent Similarity:  49.495   Percent Identity: 23.232

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-882-046-3   ..

 Align seg 1/1  to: US-08-882-046-3  from: 1  to: 4315

       8 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 24
         |||      :::   ::::::::::|||      |||      ::::::
     892 AACTGTGAGAAGGCTGAGCACGCCTGCACCTCCAACCCGTGTGCCAACGG 941

      24 rGlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluL 41
          |||   |||::::::::::|||::::::   |||   ||||||      :
     942 GGGCTCTTGCCATGAGGTGCCGTCCGGCTTCGAATGCCACTGC...CCAT 988

      41 ysGlySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyr 57
         ::|||   :::|||      |||      ::::::         :::
     989 CGGGCTGGAGCGGGCCCACCTGTGCCCTTGACATC......GATGAGTGT 1032

      58 ArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsn.. 73
           :::   ||||||      :::   |||||||:::   :::   :::
    1033 GCTTCGAACCCGTGTGCGGCCGGTGGCACCTGTGTGGACCAGGTGGACGG 1082

      74 ..................GluLysTrpLeuSerIle............. 79
                           |||:::|||::::::
    1083 CTTTGAGTGCATCTGCCCCGAGCAGTGGGTGGGGGCCACCTGCCAGCTGG 1132

      80 ..........AlaTyrGlyArgCysGlnLysIleGlyArgGlnLys 91
                 :::   |||:::|||||      |||      |||
    1133 ACGTCAACGACTGTCGCGGGCAGTGTCAGCATGGGGGCACCTGCAAG 1179

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-181-706-1

seq_documentation_block:
; Sequence 1, Application US/09181706
; Patent No. 6130068
;  GENERAL INFORMATION:
;    APPLICANT:  Melanie K. Spriggs, Michael R. Comeau,
;    APPLICANT:  Robert F. DuBose, Richard S. Johnson
;    TITLE OF INVENTION:  VIRAL ENCODED SEMAPHORIN PROTEIN
;    TITLE OF INVENTION:  RECEPTOR DNA AND POLYPEPTIDES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janis C. Henry
;      STREET:  51 University St.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  US
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/181,706
;      FILING DATE:  October 28, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/958,598 (converted to a
;      APPLICATION NUMBER:  Provisional, see below)
;      FILING DATE:  October 28, 1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  --to be assigned-- (USSN 08/958,598
;      APPLICATION NUMBER:  conversion to Provisional application)
;      FILING DATE:  October 26, 1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Henry, Janis C
;      REGISTRATION NUMBER:  34,347
;      REFERENCE/DOCKET NUMBER:  2631-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)470-4189
;      TELEFAX:  (206)233-0644
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..4707
US-09-181-706-1


alignment_scores:
            Quality:   75.00              Length:     78
              Ratio:  1.705                Gaps:      7
 Percent Similarity:  56.410   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-181-706-1   ..

 Align seg 1/1  to: US-09-181-706-1  from: 1  to: 4707

      33 AsnCysGlu..............SerHisCysAlaGluLysGlySerGl 44
         :::||||||              |||:::      |||:::|||:::||
    1296 TCCTGTGAAGAATATCTACATTTATCTAACAGCTGGGAAAGAGGTGAGGA 1345

      44 uGlySerLeuCysGlnThrGln.........ValPhePheGlyGlnTyr. 57
         |:::|||   |||:::   |||         ::::::||||||   :::
    1346 GAATTCGTGTTGCAAACTGCAATAAACATAAATCCTGTTCGGAGTGTTTA 1395

      58 ....................ArgAla.CysProCysLeuArgAsnLeuTh 67
                              |||    |||      |||::::::   ||
    1396 ACAGCCACAGACCCTCACTGCGGTTGGTGCCATTCGCTACAAAGGTGCAC 1445

      67 r............CysIleTyrSerLysAsn...GluLysTrpLeuSerI 79
         |           |||::::::|||:::|||   |||:::||||||:::|
    1446 TTTTCAAGGAGATTGTGTACATTCAGAGAACTTAGAAAACTGGCTGGATA 1495

      79 leAlaTyrGly......ArgCysGlnLysIle 87
         ||:::   |||      :::|||   ||||||
    1496 TTTCGTCTGGAGCAAAAAAGTGCCCTAAAATT 1527

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-458-791-1

seq_documentation_block:
; Sequence 1, Application US/09458791
; Patent No. 6174689
;   GENERAL INFORMATION:
;        APPLICANT: Spriggs, Melanie
;        TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
;                            RECEPTOR DNA AND POLYPEPTIDES
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Janis C. Henry
;             STREET: 51 University St.
;             CITY: Seattle
;             STATE: WA
;             COUNTRY: US
;             ZIP: 98101
;        COMPUTER READABLE FORM:
```

```
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: MS-DOS/Windows 95
;            SOFTWARE: Word for Windows 95, 7.0a
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/458,791
;            FILING DATE: 10-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/958,598
;            FILING DATE: 28-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Henry, Janis C
;            REGISTRATION NUMBER: 34,347
;            REFERENCE/DOCKET NUMBER: 2631
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206)470-4189
;            TELEFAX: (206)233-0644
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 4707 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FEATURE:
;            NAME/KEY: CDS
;            LOCATION: 1..4707
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-458-791-1


alignment_scores:
            Quality:   75.00              Length:     78
              Ratio:  1.705                Gaps:      7
 Percent Similarity:  56.410   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-458-791-1   ..

 Align seg 1/1  to: US-09-458-791-1  from: 1  to: 4707

      33 AsnCysGlu..............SerHisCysAlaGluLysGlySerGl 44
         :::||||||               |||:::      |||:::|||:::|| 
    1296 TCCTGTGAAGAATATCTACATTTATCTAACAGCTGGGAAAGAGGTGAGGA 1345

      44 uGlySerLeuCysGlnThrGln.........ValPhePheGlyGlnTyr. 57
         |:::|||   |||:::   |||         ::::::||||||   :::  
    1346 GAATTCGTGTTGCAAACTGCAATAAACATAAATCCTGTTCGGAGTGTTTA 1395

      58 ....................ArgAla.CysProCysLeuArgAsnLeuTh 67
                              |||    |||      |||::::::   ||  
    1396 ACAGCCACAGACCCTCACTGCGGTTGGTGCCATTCGCTACAAAGGTGCAC 1445

      67 r............CysIleTyrSerLysAsn...GluLysTrpLeuSerI 79
         |            |||::::::|||:::|||   |||:::|||||||:::| 
    1446 TTTTCAAGGAGATTGTGTACATTCAGAGAACTTAGAAAACTGGCTGGATA 1495

      79 leAlaTyrGly......ArgCysGlnLysIle 87
         ||:::   |||      :::|||   ||||||
    1496 TTTCGTCTGGAGCAAAAAAGTGCCCTAAAATT 1527

 seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-459-066-1

 seq_documentation_block:
 ; Sequence 1, Application US/09459066
 ; Patent No. 6187909
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Spriggs, Melanie
 ;    TITLE OF INVENTION:  VIRAL ENCODED SEMAPHORIN PROTEIN
 ;    TITLE OF INVENTION:  RECEPTOR DNA AND POLYPEPTIDES
 ;    NUMBER OF SEQUENCES:  10
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  Janis C. Henry
 ;      STREET:  51 University St.
 ;      CITY:  Seattle
 ;      STATE:  WA
 ;      COUNTRY:  US
 ;      ZIP:  98101
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  MS-DOS/Windows 95
 ;      SOFTWARE:  Word for Windows 95, 7.0a
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/09/459,066
 ;      FILING DATE:
 ;      CLASSIFICATION:
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  08/958,598
 ;      FILING DATE:
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Henry, Janis C
 ;      REGISTRATION NUMBER:  34,347
 ;      REFERENCE/DOCKET NUMBER:  2631
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  (206)470-4189
 ;      TELEFAX:  (206)233-0644
 ;  INFORMATION FOR SEQ ID NO:  1:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  4707 base pairs
 ;      TYPE:  nucleic acid
 ;      STRANDEDNESS:  double
 ;      TOPOLOGY:  linear
 ;    MOLECULE TYPE:  cDNA
 ;    HYPOTHETICAL:  NO
 ;    ANTI-SENSE:  NO
 ;    FEATURE:
 ;      NAME/KEY:  CDS
 ;      LOCATION:  1..4707
 US-09-459-066-1


 alignment_scores:
             Quality:   75.00              Length:     78
               Ratio:  1.705                Gaps:      7
  Percent Similarity:  56.410   Percent Identity: 33.333

 alignment_block:
  US-09-215-435-179_COPY_24_121 x US-09-459-066-1   ..

  Align seg 1/1  to: US-09-459-066-1  from: 1  to: 4707

       33 AsnCysGlu..............SerHisCysAlaGluLysGlySerGl 44
          :::||||||               |||:::      |||:::|||:::|| 
     1296 TCCTGTGAAGAATATCTACATTTATCTAACAGCTGGGAAAGAGGTGAGGA 1345

       44 uGlySerLeuCysGlnThrGln.........ValPhePheGlyGlnTyr. 57
          |:::|||   |||:::   |||         ::::::||||||   :::  
     1346 GAATTCGTGTTGCAAACTGCAATAAACATAAATCCTGTTCGGAGTGTTTA 1395

       58 ....................ArgAla.CysProCysLeuArgAsnLeuTh 67
                               |||    |||      |||::::::   ||  
     1396 ACAGCCACAGACCCTCACTGCGGTTGGTGCCATTCGCTACAAAGGTGCAC 1445

       67 r............CysIleTyrSerLysAsn...GluLysTrpLeuSerI 79
          |            |||::::::|||:::|||   |||:::|||||||:::| 
     1446 TTTTCAAGGAGATTGTGTACATTCAGAGAACTTAGAAAACTGGCTGGATA 1495

       79 leAlaTyrGly......ArgCysGlnLysIle 87
          ||:::   |||      :::|||   ||||||
     1496 TTTCGTCTGGAGCAAAAAAGTGCCCTAAAATT 1527

  seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-597-545-4
```

```
seq_documentation_block:
; Sequence 4, Application US/08597545
; Patent No. 5580738
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,545
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1573 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Mouse Dlk
US-08-597-545-4


alignment_scores:
            Quality:   74.00              Length:     61
              Ratio:  2.176                Gaps:      4
 Percent Similarity:  55.738   Percent Identity: 31.148

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-597-545-4   ..

 Align seg 1/1  to: US-08-597-545-4  from: 1  to: 1573

      12 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 28
         |||   :::::::|||:::           |||   :::|||   |||
     290 CTCTGTGACAAGTGTGTAACTGCCCCTGGCTGTGTCAATGGAGTCTGC.. 337

      28 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 45
          :::   |||   :::|||           ::::::   |||   :::|
     338 .AAGGAACCATGGCAGTGC.........ATCTGCAAGGATGGCTGGGACG 377

      45 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 60
         ||:::::::|||:::   :::|||             |||||||||
     378 GGAAATTCTGCGAAATAGACGTT..............CGGGCTTGCACC 412

      61 ......ProCysLeuArgAsnLeuThrCysIle 69
               ||||||   :::|||   |||||||:::
     413 TCAACCCCCTGCGCCAACAATGGAACTTGCGTG 445

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-457-135-4
```

```
seq_documentation_block:
; Sequence 4, Application US/08457135
; Patent No. 5644031
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,135
;      FILING DATE:  01-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/304/NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1573 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Mouse Dlk
US-08-457-135-4


alignment_scores:
            Quality:   74.00              Length:     61
              Ratio:  2.176                Gaps:      4
 Percent Similarity:  55.738   Percent Identity: 31.148

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-457-135-4   ..

 Align seg 1/1  to: US-08-457-135-4  from: 1  to: 1573

      12 LeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCysGl 28
         |||   :::::::|||:::           |||   :::|||   |||
     290 CTCTGTGACAAGTGTGTAACTGCCCCTGGCTGTGTCAATGGAGTCTGC.. 337

      28 nArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerGluG 45
          :::   |||   :::|||           ::::::   |||   :::|
     338 .AAGGAACCATGGCAGTGC.........ATCTGCAAGGATGGCTGGGACG 377

      45 lySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys... 60
         ||:::::::|||:::   :::|||             |||||||||
     378 GGAAATTCTGCGAAATAGACGTT..............CGGGCTTGCACC 412

      61 ......ProCysLeuArgAsnLeuThrCysIle 69
               ||||||   :::|||   |||||||:::
     413 TCAACCCCCTGCGCCAACAATGGAACTTGCGTG 445

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-9
```

```
seq_documentation_block:
; Sequence 9, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4029 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  226..2647
US-07-862-021B-9


alignment_scores:
            Quality:   73.00              Length:    103
              Ratio:  1.304                Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-07-862-021B-9   ..

 Align seg 1/1  to: US-07-862-021B-9  from: 1  to: 4029

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1852 GAAGAGACAGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCTAGCAGCTG 1901

      27 sGlnArgAlaPro.....AspAsnCysGluSerHisCysAla. 39
         |      :::           |||:::|||::::::   |||:::
    1902 CCTGGTGACTGAGTGGGGTGAGTGGGATGACTGCAGCGCCACCTGTGGAA 1951

      40 ........................GluLysGlySerGluGly 45
                                    :::   :::::::|||
    1952 TGGGCATGAAGAAGCGGCACCGCATGGTCAAGATGAGCCCCGCGGACGGC 2001

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::             |||
    2002 TCCATGTGCAAGGCGGAGACTTCGCAGGCGGAGAAATGCATGATGCCTGA 2051

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp.....L 77
            :::::::::   |||:::   |||      ::::::|||
    2052 GTGCCATACCATCCCGTGCTTGCTGTCTCCTTGGTCCGAGTGGAGCGACT 2101

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||:::::::   |||:::   ::::::      ||||||::::::
    2102 GTAGCGTGACCTGTGGGAAGGGCATGCGGACG...CGCCAGCGGATGCTC 2148

      94 LysLysMet 96
         |||::::::
    2149 AAGTCTCTG 2157

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-9

seq_documentation_block:
; Sequence 9, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  226..2647
US-08-313-288B-9


alignment_scores:
            Quality:   73.00              Length:    103
              Ratio:  1.304                Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-313-288B-9   ..

 Align seg 1/1  to: US-08-313-288B-9  from: 1  to: 4029

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1852 GAAGAGACAGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCTAGCAGCTG 1901

      27 sGlnArgAlaPro.....AspAsnCysGluSerHisCysAla. 39
```

```
        |      :::               |||:::|||::::::   |||:::
    1902 CCTGGTGACTGAGTGGGGTGAGTGGGATGACTGCAGCGCCACCTGTGGAA 1951

      40 ..............................GluLysGlySerGluGly 45
                                      :::   :::::::|||
    1952 TGGGCATGAAGAAGCGGCACCGCATGGTCAAGATGAGCCCCGCGGACGGC 2001

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::            |||
    2002 TCCATGTGCAAGGCGGAGACTTCGCAGGCGGAGAAATGCATGATGCCTGA 2051

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
           :::::::::   |||:::  |||      ::::::|||
    2052 GTGCCATACCATCCCGTGCTTGCTGTCTCCTTGGTCCGAGTGGAGCGACT 2101

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::  ::::::      ||||||::::::
    2102 GTAGCGTGACCTGTGGGAAGGGCATGCGGACG...CGCCAGCGGATGCTC 2148

      94 LysLysMet 96
         |||::::::
    2149 AAGTCTCTG 2157
```

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-9

```
seq_documentation_block:
; Sequence 9, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4029 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  226..2647
PCT-US93-03164-9
```

```
alignment_scores:
           Quality:   73.00            Length:    103
             Ratio:  1.304              Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x PCT-US93-03164-9   ..

 Align seg 1/1  to: PCT-US93-03164-9  from: 1  to: 4029

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||:::::::|||:::   :::   ||
    1852 GAAGAGACAGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCTAGCAGCTG 1901

      27 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 39
         |      :::               |||:::|||::::::   |||:::
    1902 CCTGGTGACTGAGTGGGGTGAGTGGGATGACTGCAGCGCCACCTGTGGAA 1951

      40 ..............................GluLysGlySerGluGly 45
                                      :::   ::::::|||
    1952 TGGGCATGAAGAAGCGGCACCGCATGGTCAAGATGAGCCCCGCGGACGGC 2001

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::            |||
    2002 TCCATGTGCAAGGCGGAGACTTCGCAGGCGGAGAAATGCATGATGCCTGA 2051

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
           :::::::::   |||:::  |||      ::::::|||
    2052 GTGCCATACCATCCCGTGCTTGCTGTCTCCTTGGTCCGAGTGGAGCGACT 2101

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::  ::::::      ||||||::::::
    2102 GTAGCGTGACCTGTGGGAAGGGCATGCGGACG...CGCCAGCGGATGCTC 2148

      94 LysLysMet 96
         |||::::::
    2149 AAGTCTCTG 2157
```

```
OM of: US-09-215-435-179_COPY_24_121 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78 +  547.00 1126.42  2.1e-53   542
/cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61 +  547.00 1125.99  2.2e-53   565
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73 +  547.00 1124.11  2.9e-53   680
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73 +  547.00 1124.11  2.9e-53   680
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817 +  542.00 1116.18  7.9e-53   533
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817 +  542.00 1116.18  7.9e-53   53
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481-520 +  532.00 1096.25  1.0e-51   488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481A-520 +  532.00 1096.25  1.0e-51   488
/cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481B-520 +  532.00 1096.25  1.0e-51   488
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44 +  500.00 1031.13  4.3e-48   421
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-48 +  497.00 1023.92  1.1e-47   463
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-694 +  497.00 1023.59  1.1e-47   478
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-31337-1 +  497.00 1022.68  1.3e-47   523
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-443-634-1 +  497.00 1022.68  1.3e-47   523
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-713-067-1 +  497.00 1022.68  1.3e-47   523
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-992-332-1466 +  383.00  792.32  8.6e-35   263
/cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-033-364-14 +  383.00  792.32  8.6e-35   263
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-192 +  343.00  643.02  1.8e-26   171
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-212-664-219 +  343.00  643.02  1.8e-26   171
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-731-872-89 +  264.50  538.77  1.1e-20   514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-5913 +  264.50  538.77  1.1e-20   514
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-876-997-89 +  264.50  538.77  1.1e-20   514
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-1374 +  264.50  538.77  1.1e-20   51
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-108 +  264.50  538.77  1.1e-20   514
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-108 +  264.50  538.77  1.1e-20   514
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-5913 +  264.50  538.77  1.1e-20   51
/cgn1_7/ptodata/1/pna/US095A_COMB.seq:US-09-528-409-86528 +  188.00  380.94  7.1e-12   4
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-933-524-86528 +  188.00  380.94  7.1e-12   44
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-274-861-9011 +  135.00  271.78  8.5e-06   397
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-915-738-9011 +  135.00  271.78  8.5e-06   397
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-284-697A-15 +  135.00  268.97  1.2e-05   523
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-654-617-289898 +  135.00  268.00  1.4e-05
/cgn1_7/ptodata/1/pna/US096E_COMB.seq:US-09-684-016-289898 +  135.00  268.00  1.4e-05
/cgn1_7/ptodata/1/pna/US6021_COMB.seq:US-60-213-360-121 +  135.00  267.74  1.4e-05   590
/cgn1_7/ptodata/1/pna/US6032_COMB.seq:US-60-324-185-6544 +  135.00  267.74  1.4e-05   59
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-7251 +  135.00  267.57  1.5e-05   60
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-122 +  135.00  267.35  1.5e-05   613
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-647-801-16 +  135.00  266.66  1.6e-05   656
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-05989-156 +  135.00  266.51  1.7e-05   666
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-925-297-156 +  135.00  266.51  1.7e-05   666
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-8817 +  135.00  265.37  1.9e-05
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-14399 +  131.00  252.47  0.0001
/cgn1_7/ptodata/1/pna/US6032_COMB.seq:US-60-324-185-6539 +  131.00  251.83  0.0001
/cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-705-426-361 +  131.00  251.81  0.0001
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-177-732-8 +  131.00  251.81  0.0001    12

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-78

seq_documentation_block:
; Sequence 78, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 78
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 16..378
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 16..84
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 9.8
;   OTHER INFORMATION: seq FLLFFFLFLLLTRG/SL
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 502..507
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 522..542
US-09-215-435-78


alignment_scores:
            Quality:  547.00              Length:     98
              Ratio:  5.582                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-215-435-78   ..

 Align seg 1/1  to: US-09-215-435-78  from: 1  to: 542

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 TCACTTTCTCCAACAAAATATAACCTTTTGGAGCTCAAGGAGTCTTGCAT 134

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     135 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 184

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 234

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      235 CAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 284

       67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      285 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 334

       84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
          ||||||||||||||||||||||||||||||||||||||||||
      335 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 378

seq_name: /cgn1_7/ptodata/1/pna/US6009_COMB.seq:US-60-096-116-61

seq_documentation_block:
; Sequence 61, Application US/60096116
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  95
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/096,116
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.038PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  565 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  16..84
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLTRG/SL
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  502..507
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  529..542
;      IDENTIFICATION METHOD:  blastn2
;    FEATURE:
;      NAME/KEY:  vrt
;      LOCATION:  complement(33..74)
;      IDENTIFICATION METHOD:  fasta
;      OTHER INFORMATION:  identity 88
;      OTHER INFORMATION:  region 21722..21764
;      OTHER INFORMATION:  id AF048728
;    FEATURE:
;      NAME/KEY:  vrt
;      LOCATION:  511..545
;      IDENTIFICATION METHOD:  fasta
;      OTHER INFORMATION:  identity 83
;      OTHER INFORMATION:  region 4627..4663
;      OTHER INFORMATION:  id AF048728
US-60-096-116-61


alignment_scores:
            Quality:  547.00              Length:     98
              Ratio:  5.582                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-60-096-116-61    ..

 Align seg 1/1  to: US-60-096-116-61  from: 1  to: 565

        1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       85 TCACTTTCTCCAACAAAATATAACCTTTTGGAGCTCAAGGAGTCTTGCAT 134

       17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      135 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 184

       34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      185 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 234

       51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      235 CAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 284

       67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      285 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 334

       84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
          ||||||||||||||||||||||||||||||||||||||||||
      335 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 378

seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-17130-73

seq_documentation_block:
; Sequence 73, Application PC/TUS9917130
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/17130
;  CURRENT FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
```

```
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-17130-73


alignment_scores:
           Quality:  547.00            Length:     98
             Ratio:  5.582               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x PCT-US99-17130-73   ..

 Align seg 1/1  to: PCT-US99-17130-73  from: 1  to: 680

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 251

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 301

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 351

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 CAGGTGTTCTTTGGCCAGTATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 401

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 451

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         |||||||||||||||||||||||||||||||||||||||||||
     452 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 495

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-489-847-73

seq_documentation_block:
; Sequence 73, Application US/09489847
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031P1
;  CURRENT APPLICATION NUMBER: US/09/489,847
;  CURRENT FILING DATE: 2000-01-24
;  EARLIER APPLICATION NUMBER: PCT/US99/17130
;  EARLIER FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 376
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
;   LENGTH: 680
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-489-847-73


alignment_scores:
           Quality:  547.00            Length:     98
             Ratio:  5.582               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-489-847-73   ..

 Align seg 1/1  to: US-09-489-847-73  from: 1  to: 680

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 251

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 301

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 351

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 CAGGTGTTCTTTGGCCAGTATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 401

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 451

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         |||||||||||||||||||||||||||||||||||||||||||
     452 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 495

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4817

seq_documentation_block:
; Sequence 4817, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
```

```
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 4817
;    LENGTH: 533
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 99..461
;    NAME/KEY: sig_peptide
;    LOCATION: 99..167
;    OTHER INFORMATION: Von Heijne matrix
;    OTHER INFORMATION: score 9.80000019073486
;    OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
US-09-834-366-4817


alignment_scores:
            Quality:  542.00              Length:     98
              Ratio:  5.588                 Gaps:      0
 Percent Similarity:  98.980   Percent Identity: 98.980

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-834-366-4817    ..

 Align seq 1/1  to: US-09-834-366-4817  from: 1  to: 533

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     168 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 217

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 267

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         |||||||||||||||||||||||||||||||||   |||||||||
     268 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCKGTGTCAAACG 317

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     318 CAGGTGTTCTTTGGCCARTATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 367

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     368 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 417

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         ||||||||||||||||||||||||||||||||||||||||||
     418 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAARAAAATGTTCTTC 461

 seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4817

 seq_documentation_block:
 ; Sequence 4817, Application US/60197873
 ; GENERAL INFORMATION:
 ;  APPLICANT:  Bejanin, Stephane
 ;  APPLICANT:  Tanaka, Hiroaki
 ;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
 ;  APPLICANT:  Jobert, Severin
 ;  APPLICANT:  Giordano, Jean-Yves
 ;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
 ;  FILE REFERENCE:  81.US1.PRO
 ;  CURRENT APPLICATION NUMBER: US/60/197,873
 ;  CURRENT FILING DATE:  2000-04-18
 ;  NUMBER OF SEQ ID NOS:  52153
 ;  SOFTWARE:  Patent.pm
 ; SEQ ID NO 4817
 ;    LENGTH: 533
 ;    TYPE: DNA
 ;    ORGANISM: Homo sapiens
 ;    FEATURE:
 ;    NAME/KEY: CDS
 ;    LOCATION: 99..461
 ;    NAME/KEY: sig_peptide
 ;    LOCATION: 99..167
 ;    OTHER INFORMATION: Von Heijne matrix
 ;    OTHER INFORMATION: score 9.80000019073486
 ;    OTHER INFORMATION: seq FLLFFFLFLLTRG/SL
 US-60-197-873-4817


 alignment_scores:
             Quality:  542.00              Length:     98
               Ratio:  5.588                 Gaps:      0
  Percent Similarity:  98.980   Percent Identity: 98.980

 alignment_block:
  US-09-215-435-179_COPY_24_121 x US-60-197-873-4817    ..

  Align seq 1/1  to: US-60-197-873-4817  from: 1  to: 533

        1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      168 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 217

       17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      218 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 267

       34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
          |||||||||||||||||||||||||||||||||   |||||||||
      268 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCKGTGTCAAACG 317

       51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      318 CAGGTGTTCTTTGGCCARTATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 367

       67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      368 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 417

       84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
          ||||||||||||||||||||||||||||||||||||||||||
      418 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAARAAAATGTTCTTC 461

  seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481-520

  seq_documentation_block:
  ; Sequence 520, Application US/08196481
  ;  GENERAL INFORMATION:
  ;    APPLICANT:  Haseltine, William
  ;    APPLICANT:  Rosen, Craig
  ;    APPLICANT:  Ruben, Steve
  ;    APPLICANT:  Dillon, Patrick
  ;    APPLICANT:  Li, Haodong
  ;    TITLE OF INVENTION:  Human Genes, Sequences, and Expression
  ;    TITLE OF INVENTION:  Products
  ;    NUMBER OF SEQUENCES:  1719
  ;    CORRESPONDENCE ADDRESS:
  ;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart, &
  ;      ADDRESSEE:  Olstein
  ;      STREET:  6 Becker Farm Road
  ;      CITY:  Roseland
  ;      STATE:  New Jersey
  ;      COUNTRY:  USA
  ;      ZIP:  07068
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 5.0
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/196,481
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  325800-92
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 994-1700
;      TELEFAX:  (201) 994-1744
;  INFORMATION FOR SEQ ID NO:  520:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-196-481-520


alignment_scores:
            Quality:  532.00              Length:     98
              Ratio:  5.485                 Gaps:      0
 Percent Similarity:  98.980   Percent Identity: 96.939

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-196-481-520   ..

 Align seg 1/1  to: US-08-196-481-520  from: 1  to: 488

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      56 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 105

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     106 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 155

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     156 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 205

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 CAGGTGTTCTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 255

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 305

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         ||||||||||| ||||||||||:::|||:::|||||||||
     306 GTCAGAAAATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349

seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481A-520

seq_documentation_block:
; Sequence 520, Application US/08196481A
;  GENERAL INFORMATION:
;    APPLICANT:  Haseltine, William A.
;    APPLICANT:  Rosen, Craig A.
;    APPLICANT:  Ruben, Steve
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Li, Haodong
;    TITLE OF INVENTION:  Human Genes, Sequences and Expression                Products
;    NUMBER OF SEQUENCES:  1721
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/196,481A
;      FILING DATE:  February 15, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenley K. Hoover
;      REGISTRATION NUMBER:  40,302
;      REFERENCE/DOCKET NUMBER:  PO-07
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8439
;  INFORMATION FOR SEQ ID NO:  520:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-196-481A-520


alignment_scores:
            Quality:  532.00              Length:     98
              Ratio:  5.485                 Gaps:      0
 Percent Similarity:  98.980   Percent Identity: 96.939

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-196-481A-520   ..

 Align seg 1/1  to: US-08-196-481A-520  from: 1  to: 488

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      56 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 105

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     106 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 155

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     156 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 205

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 CAGGTGTTCTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 255

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 305

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         ||||||||||| ||||||||||:::|||:::|||||||||
     306 GTCAGAAAATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349

seq_name: /cgn1_7/ptodata/1/pna/US081_COMB.seq:US-08-196-481B-520

seq_documentation_block:
```

```
; Sequence 520, Application US/08196481B
; GENERAL INFORMATION:
;  APPLICANT:  Rosen, et. al.
;  TITLE OF INVENTION:  Human Genes, Sequences, and Expression Products 7
;  FILE REFERENCE:  PO07
;  CURRENT APPLICATION NUMBER: US/08/196,481B
;  CURRENT FILING DATE:  1994-02-15
;  NUMBER OF SEQ ID NOS:  1721
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 520
;   LENGTH: 488
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (220)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (317)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (354)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (356)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (358)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (366)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (370)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (378)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (425)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (432)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (433)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (434)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (437)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (439)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (440)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (449)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (455)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (457)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (458)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (463)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (482)
;   OTHER INFORMATION: n equals a,t,g, or c
US-08-196-481B-520


alignment_scores:
            Quality:  532.00              Length:     98
              Ratio:  5.485                 Gaps:      0
 Percent Similarity:  98.980   Percent Identity: 96.939

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-196-481B-520   ..

 Align seg 1/1  to: US-08-196-481B-520  from: 1  to: 488

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      56 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCAT 105

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     106 CCGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 155

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     156 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 205

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     206 CAGGTGTTCTTTGGNCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 255

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     256 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 305

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
         ||||||||||||   |||||||||||:::|||:::|||||||||
     306 GTCAGAAAATTNGAAGGCAGAAGTTGGTTAAGGAAATGTTCTTT 349

 seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-279-44

 seq_documentation_block:
 ; Sequence 44, Application US/08905279
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Genset SA
 ;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
 ;    TITLE OF INVENTION:  EXPRESSED IN TESTIS AND OTHER TISSUES
 ;    NUMBER OF SEQUENCES:  503
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
 ;      STREET:  501 West Broadway
 ;      CITY:  San Diego
 ;      STATE:  California
 ;      COUNTRY:  USA
 ;      ZIP:  92101-3505
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy Disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  Win95
 ;      SOFTWARE:  Word
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/905,279
 ;      FILING DATE:
 ;      CLASSIFICATION:  536
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Israelsen, Ned A.
 ;      REGISTRATION NUMBER:  29,655
 ;      REFERENCE/DOCKET NUMBER:  GENSET.011A
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  (619) 235-8550
 ;      TELEFAX:  (619) 235-0176
```

```
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  421 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Testis
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  62..130
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 9.8
;      OTHER INFORMATION:  seq FLLFFFLFLLTRG/SL
US-08-905-279-44


alignment_scores:
            Quality:  500.00              Length:     95
              Ratio:  5.435                 Gaps:      0
 Percent Similarity:  96.842   Percent Identity: 94.737

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-08-905-279-44    ..

 Align seg 1/1  to: US-08-905-279-44  from: 1  to: 421

       1 SerLeuSerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIl 17
         ||||||||||||||||||||||||||||||||||||||||||||||:::
     131 TCACTTTCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGKSKKGCAT 180

      17 eArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 SGGGAACCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATT 230

      34 ysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGlnThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 GCGAGTCGCACTGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACG 280

      51 GlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CAGGTGTTCTTTGGCCAATATAGAGCGTGTCCCTGCCTGCGGAACCTGAC 330

      67 rCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgC 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 TTGTATATATTCAAAGAATGAGAAATGGCTTAGCATCGCCTATGGCCGTT 380

      84 ysGlnLysIleGlyArgGlnLysLeuAlaLysLys 95
         ||||||||||||||||||||||||||||||:::|||
     381 GTCAGAAAATTGGAAGGCAGAAGTTGGCTAGRAAA 415

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-48

seq_documentation_block:
; Sequence 48, Application US/60184698
; GENERAL INFORMATION:
;  APPLICANT: Hodgson, David M.                Jones, Anissa L.
;  APPLICANT: Lincoln, Stephen E.          Yu, Jimmy Y.
;  APPLICANT: Russo, Frank D.                  Greenawalt, Lila B.
;  APPLICANT: Spiro, Peter A.                  Panzer, Scott R.
;  APPLICANT: Banville, Steve C.           Roseberry, Ann M.
;  APPLICANT: Bratcher, Shawn R.           Wright, Rachel J.
;  APPLICANT: Dufour, Gerard E.                Chen, Wensheng
;  APPLICANT: Cohen, Howard J.                 Liu, Tommy
;  APPLICANT: Rosen, Bruce                 Yap, Pierre E.
;  APPLICANT: Shah, Purvi                      Stockdreher, Theresa K.
;  APPLICANT: Chalup, Michael S.           Amshey, Stefan
;  APPLICANT: Hillman, Jennifer L.             Fong, Willy Tuen
;  TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
;  FILE REFERENCE: PT-0115 P
;  CURRENT APPLICATION NUMBER: US/60/184,698
;  CURRENT FILING DATE: 2000-02-24
;  NUMBER OF SEQ ID NOS: 1132
;  SOFTWARE: PERL Program
; SEQ ID NO 48
;   LENGTH: 463
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 474641.1
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 347-368
;   OTHER INFORMATION: a, t, c, g, or other
US-60-184-698-48


alignment_scores:
            Quality:  497.00              Length:     88
              Ratio:  5.648                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-60-184-698-48    ..

 Align seg 1/1  to: US-60-184-698-48  from: 1  to: 463

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      61 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 110

      27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 160

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 210

      61 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 77
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 260

      77 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 310

      94 ysLysMetPhePhe 98
         |||||||||||||
     311 AGAAAATGTTCTTC 324

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-184-698-694

seq_documentation_block:
; Sequence 694, Application US/60184698
; GENERAL INFORMATION:
;  APPLICANT: Hodgson, David M.                Jones, Anissa L.
;  APPLICANT: Lincoln, Stephen E.          Yu, Jimmy Y.
;  APPLICANT: Russo, Frank D.                  Greenawalt, Lila B.
;  APPLICANT: Spiro, Peter A.                  Panzer, Scott R.
;  APPLICANT: Banville, Steve C.           Roseberry, Ann M.
;  APPLICANT: Bratcher, Shawn R.           Wright, Rachel J.
;  APPLICANT: Dufour, Gerard E.                Chen, Wensheng
;  APPLICANT: Cohen, Howard J.                 Liu, Tommy
;  APPLICANT: Rosen, Bruce                 Yap, Pierre E.
;  APPLICANT: Shah, Purvi                      Stockdreher, Theresa K.
;  APPLICANT: Chalup, Michael S.           Amshey, Stefan
;  APPLICANT: Hillman, Jennifer L.             Fong, Willy Tuen
;  TITLE OF INVENTION: BIOCHEMICAL PATHWAY MOLECULES
;  FILE REFERENCE: PT-0115 P
;  CURRENT APPLICATION NUMBER: US/60/184,698
;  CURRENT FILING DATE: 2000-02-24
;  NUMBER OF SEQ ID NOS: 1132
```

```
;  SOFTWARE: PERL Program
; SEQ ID NO 694
;   LENGTH: 478
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 474641.1.i
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 347-368
;   OTHER INFORMATION: a, t, c, g, or other
US-60-184-698-694


alignment_scores:
            Quality:  497.00              Length:     88
              Ratio:  5.648                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-60-184-698-694    ..

 Align seg 1/1  to: US-60-184-698-694  from: 1  to: 478

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      61 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 110

      27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     111 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 160

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     161 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 210

      61 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 77
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 260

      77 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 94
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 310

      94 ysLysMetPhePhe 98
         |||||||||||||||
     311 AGAAAATGTTCTTC 324

 seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US00-31337-1

 seq_documentation_block:
 ; Sequence 1, Application PC/TUS0031337
 ; GENERAL INFORMATION:
 ;  APPLICANT:  ZymoGenetics, Inc.
 ;  TITLE OF INVENTION:  EDUCATIONAL KIT AND METHOD USING
 ;  TITLE OF INVENTION:  COLIPASE-LIKE PROTEIN-1
 ;  FILE REFERENCE:  99-85PC
 ;  CURRENT APPLICATION NUMBER: PCT/US00/31337
 ;  CURRENT FILING DATE:  2000-11-15
 ;  PRIOR APPLICATION NUMBER:  60/166,626
 ;  PRIOR FILING DATE:  1999-11-19
 ;  NUMBER OF SEQ ID NOS:  11
 ;  SOFTWARE:  FastSEQ for Windows Version 3.0
 ; SEQ ID NO 1
 ;   LENGTH: 523
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: (1)...(354)
 PCT-US00-31337-1


alignment_scores:
            Quality:  497.00              Length:     88
              Ratio:  5.648                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x PCT-US00-31337-1    ..

 Align seg 1/1  to: PCT-US00-31337-1  from: 1  to: 523

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

      27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      61 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 77
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

      77 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 94
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

      94 ysLysMetPhePhe 98
         |||||||||||||||
     341 AGAAAATGTTCTTC 354

 seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-443-634-1

 seq_documentation_block:
 ; Sequence 1, Application US/09443634
 ; GENERAL INFORMATION:
 ;  APPLICANT: Holloway, James L.
 ;  TITLE OF INVENTION: Colipase-Like Protein - 1
 ;  FILE REFERENCE: 99-85
 ;  CURRENT APPLICATION NUMBER: US/09/443,634
 ;  CURRENT FILING DATE: 1999-11-19
 ;  NUMBER OF SEQ ID NOS: 11
 ;  SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 1
 ;   LENGTH: 523
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: (1)...(354)
 US-09-443-634-1


 alignment_scores:
             Quality:  497.00              Length:     88
               Ratio:  5.648                Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-179_COPY_24_121 x US-09-443-634-1    ..

  Align seg 1/1  to: US-09-443-634-1  from: 1  to: 523

       11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

       27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      61 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 77
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

      77 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

      94 ysLysMetPhePhe 98
         ||||||||||||||
     341 AGAAAATGTTCTTC 354

seq_name: /cgn1_7/ptodata/1/pna/US097A_COMB.seq:US-09-713-067-1

seq_documentation_block:
; Sequence 1, Application US/09713067
; GENERAL INFORMATION:
;  APPLICANT:  Holloway, James L.
;  TITLE OF INVENTION:  EDUCATIONAL KIT AND METHOD USING
;  TITLE OF INVENTION:  COLIPASE-LIKE PROTEIN-1
;  FILE REFERENCE:  99-85US
;  CURRENT APPLICATION NUMBER: US/09/713,067
;  CURRENT FILING DATE:  2000-11-15
;  PRIOR APPLICATION NUMBER:  60/166,626
;  PRIOR FILING DATE:  1999-11-19
;  NUMBER OF SEQ ID NOS:  11
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 523
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(354)
US-09-713-067-1


alignment_scores:
           Quality:  497.00            Length:     88
             Ratio:  5.648              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-713-067-1   ..

 Align seg 1/1  to: US-09-713-067-1  from: 1  to: 523

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 140

      27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     141 CCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAAGGGGTCCG 190

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCys 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     191 AGGGCAGTCTGTGTCAAACGCAGGTGTTCTTTGGCCAATATAGAGCGTGT 240

      61 ProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrpLe 77
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCCTGCCTGCGGAACCTGACTTGTATATATTCAAAGAATGAGAAATGGCT 290

      77 uSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAlaL 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     291 TAGCATCGCCTATGGCCGTTGTCAGAAAATTGGAAGGCAGAAGTTGGCTA 340

      94 ysLysMetPhePhe 98
         ||||||||||||||
     341 AGAAAATGTTCTTC 354
```

```
OM of: US-09-215-435-179_COPY_24_121 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-9 +   74.50  136.21  2.26      308
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-7 +   74.50  131.88  3.94      452
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-206 +   74.00  138.06  1.78      2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-184 +   74.00  135.07  2.62      3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-177 +   74.00  132.10  3.83      3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-353 +   74.00  129.88  5.09      4
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5020 +   71.50  151.57  0.3152
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5019 +   71.50  150.70  0.3524
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-2381 -   71.50  134.08  2.97
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-702 +   71.00  114.47  36.72     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-189-455 +   70.00  133.67  3.13      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-192-455 +   70.00  133.67  3.13      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-697-455 +   70.00  133.67  3.13      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-824-455 +   70.00  133.67  3.13      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-298-88 +   70.00  131.76  4.00      17
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-1085 +   70.00  131.76  4.00
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1145 +   70.00  126.74  7.61
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-786 +   69.00  131.15  4.32      1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-619-252-1 +   69.00  115.07  34.02     637
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8642 +   66.50  139.89  1.41
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-8641 +   66.00  138.77  1.63
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-971-392-43 -   66.00  125.02  9.49      14
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-862 +   66.00  109.21  72.12     5
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-351 +   65.50  100.71  214.41    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2006 +   65.50  100.71  214.41
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2007 +   65.50  100.71  214.41
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-973-278-680 +   65.00  125.38  9.07      1
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-560 +   65.00  118.02  23.30     2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-1210 +   64.50  134.85  2.69
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-886-055-28 -   64.50  125.18  9.30      10
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-518 +   64.50  113.84  39.81     2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-858-5 +   64.00   96.99  345.92    111
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-332 -   64.00   93.01  576.29    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-472 +   64.00   93.01  576.29    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-473 +   64.00   93.01  576.29    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2390 -   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2391 -   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2392 +   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2393 +   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2394 +   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2395 +   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2396 -   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2397 +   64.00   93.01  576.29
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-742 +   63.50  122.66  12.85
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4416 +   63.50  110.44  61.63

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-9

seq_documentation_block:
; Sequence 9, Application US/09703559
; GENERAL INFORMATION:
;  APPLICANT: Emory University
;  TITLE OF INVENTION: Insulin-Responsive Sequence DNA Binding Protein-1 and Methods
;  TITLE OF INVENTION:  Insulin-Responsive Genes
;  FILE REFERENCE: E056 1010
;  CURRENT APPLICATION NUMBER: US/09/703,559
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 3087
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1)..(3084)
US-09-703-559-9


alignment_scores:
            Quality:   74.50              Length:     63
              Ratio:  2.328                Gaps:      5
 Percent Similarity:  50.794   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-703-559-9   ..

 Align seg 1/1  to: US-09-703-559-9  from: 1  to: 3087

      21 AspCysGluThrGly..............CysCysGlnArgAlaProAs 32
         :::|||      |||               |||   :::|||:::|||||
     205 GAGTGCCCCGAAGGCTTCATGGGCCTGGACTGCAGGGAGAGAGTCCCCGA 254

      32 pAsnCysGlu............SerHisCysAlaGluLysGlySerGluG 45
         |:::||||||            ::::::|||        |||::::::
     255 TGACTGTGAGTGCCGCAACGGAGGCAGATGCCTG......GGCGCCAACA 298

      45 lySerLeuCysGlnThrGlnVal...PhePheGly.............. 55
         :::|||||||||      :::   |||||||||
     299 CCACCCTCTGCCAGTGCCCCCTGGGATTCTTTGGGCTTCTCTGTGAATTT 348

      56 GlnTyrArgAlaCysProCysLeuArgAsnLeuThrCys 68
         :::      |||   ||||||      |||      |||
     349 GAAATCACAGCCATGCCCTGCAACATGAACACACAGTGC 387

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-703-559-7

seq_documentation_block:
; Sequence 7, Application US/09703559
; GENERAL INFORMATION:
;  APPLICANT: Emory University
;  TITLE OF INVENTION: Insulin-Responsive Sequence DNA Binding Protein-1 and Methods
;  TITLE OF INVENTION:  Insulin-Responsive Genes
;  FILE REFERENCE: E056 1010
;  CURRENT APPLICATION NUMBER: US/09/703,559
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 4524
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: 5'UTR
;   LOCATION: (1)..(58)
;   NAME/KEY: exon
;   LOCATION: (59)..(3082)
;   NAME/KEY: 3'UTR
;   LOCATION: (3083)..(4524)
US-09-703-559-7


alignment_scores:
            Quality:   74.50              Length:     63
              Ratio:  2.328                Gaps:      5
 Percent Similarity:  50.794   Percent Identity: 33.333

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-703-559-7   ..

 Align seg 1/1  to: US-09-703-559-7  from: 1  to: 4524

      21 AspCysGluThrGly..............CysCysGlnArgAlaProAs 32
         :::|||      |||                |||  :::|||:::|||||
     263 GAGTGCCCCGAAGGCTTCATGGGCCTGGACTGCAGGGAGAGAGTCCCCGA 312

      32 pAsnCysGlu............SerHisCysAlaGluLysGlySerGluG 45
         |:::||||||           :::::::|||        |||::::::
     313 TGACTGTGAGTGCCGCAACGGAGGCAGATGCCTG......GGCGCCAACA 356

      45 lySerLeuCysGlnThrGlnVal...PhePheGly.............. 55
           :::||||||||||      :::   |||||||||
     357 CCACCCTCTGCCAGTGCCCCCTGGGATTCTTTGGGCTTCTCTGTGAATTT 406

      56 GlnTyrArgAlaCysProCysLeuArgAsnLeuThrCys 68
         :::      |||   ||||||      |||      |||
     407 GAAATCACAGCCATGCCCTGCAACATGAACACACAGTGC 445

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-206

seq_documentation_block:
; Sequence 206, Application US/09970966
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206
;   LENGTH: 2364
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-206


alignment_scores:
            Quality:   74.00              Length:    103
              Ratio:  1.321                Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-970-966-206   ..

 Align seg 1/1  to: US-09-970-966-206  from: 1  to: 2364

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||:::::::|||:::   :::   ||
    1567 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1616

      27 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 39
         |     :::                |||:::|||::::::   |||:::
    1617 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1666

      40 ..............................GluLysGlySerGluGly 45
                                      :::   :::::::|||
    1667 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1716

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::      ::::::          |||
    1717 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1766

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
           :::::::::   |||:::   |||       :::::::|||
    1767 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1816

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
         |||::::::   |||:::  ::::::      ||||||::::::
    1817 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1863

      94 LysLysMet 96
         |||::::::
    1864 AAGTCTCTG 1872

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-184

seq_documentation_block:
; Sequence 184, Application US/09970966
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184
;   LENGTH: 3079
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-184


alignment_scores:
            Quality:   74.00              Length:    103
              Ratio:  1.321                Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-970-966-184   ..

 Align seg 1/1  to: US-09-970-966-184  from: 1  to: 3079

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||:::::::|||:::   :::   ||
    1651 GAGGAAATGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1700

      27 sGlnArgAlaPro............AspAsnCysGluSerHisCysAla. 39
         |     :::                |||:::|||::::::   |||:::
    1701 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1750

      40 ..............................GluLysGlySerGluGly 45
                                      :::   :::::::|||
    1751 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1800

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::      ::::::          |||
    1801 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1850

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
```

```
          :::::::::   |||:::   |||      ::::::|||
    1851 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1900

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
         |||::::::   |||:::   ::::::      ||||||::::::
    1901 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1947

      94 LysLysMet 96
         |||::::::
    1948 AAGTCTCTG 1956

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-970-966-177

seq_documentation_block:
; Sequence 177, Application US/09970966
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 177
;   LENGTH: 3999
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-177


alignment_scores:
            Quality:   74.00             Length:    103
              Ratio:  1.321               Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-970-966-177   ..

 Align seg 1/1  to: US-09-970-966-177  from: 1  to: 3999

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1078 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1127

      27 sGlnArgAlaPro.....AspAsnCysGluSerHisCysAla. 39
         |      :::           |||:::|||::::::   |||:::
    1128 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1177

      40 ............................GluLysGlySerGluGly 45
                                        :::   ::::::|||
    1178 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1227

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::           |||
    1228 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1277

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
         :::::::::   |||:::   |||      ::::::|||
    1278 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1327

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::   ::::::      ||||||::::::
    1328 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1374

      94 LysLysMet 96
         |||::::::
    1375 AAGTCTCTG 1383

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-353

seq_documentation_block:
; Sequence 353, Application US/09976594
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 353
;   LENGTH: 4866
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 415378.3
US-09-976-594-353


alignment_scores:
            Quality:   74.00             Length:    103
              Ratio:  1.321               Gaps:      4
 Percent Similarity:  54.369   Percent Identity: 20.388

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-976-594-353   ..

 Align seg 1/1  to: US-09-976-594-353  from: 1  to: 4866

      11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
         |||      |||:::|||      |||::::::|||:::   :::   ||
    1655 GAGGAAACGGAGAAGTGCACGGTCAACGAGGAGTGCTCTCCCAGCAGCTG 1704

      27 sGlnArgAlaPro.....AspAsnCysGluSerHisCysAla. 39
         |      :::           |||:::|||::::::   |||:::
    1705 CCTGATGACCGAGTGGGGCGAGTGGGACGAGTGCAGCGCCACCTGCGGCA 1754

      40 ............................GluLysGlySerGluGly 45
                                        :::   ::::::|||
    1755 TGGGCATGAAGAAGCGGCACCGCATGATCAAGATGAACCCCGCAGATGGC 1804

      46 SerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAlaCysProCy 62
         |||:::|||:::::::::::::      ::::::           |||
    1805 TCCATGTGCAAAGCCGAGACATCACAGGCAGAGAAGTGCATGATGCCAGA 1854

      62 sLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTrp......L 77
         :::::::::   |||:::   |||      ::::::|||
    1855 GTGCCACACCATCCCATGCTTGCTGTCCCCATGGTCCGAGTGGAGTGACT 1904

      77 euSerIleAlaTyrGlyArgCysGlnLysIleGlyArgGlnLysLeuAla 93
           |||::::::   |||:::   ::::::      ||||||::::::
    1905 GCAGCGTGACCTGCGGGAAGGGCATGCGAACC...CGACAGCGGATGCTC 1951

      94 LysLysMet 96
         |||::::::
    1952 AAGTCTCTG 1960

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5020

seq_documentation_block:
; Sequence 5020, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
```

```
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5020
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-340-5020


alignment_scores:
            Quality:   71.50              Length:     84
              Ratio:  1.663                Gaps:      3
 Percent Similarity:  51.190   Percent Identity: 25.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-922-340-5020   ..

 Align seg 1/1  to: US-09-922-340-5020  from: 1  to: 427

      26 CysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGl 42
         ||||||   :::::::|||   :::          ||||||   :::::::||
      44 TGCTGTCCACATACCCCAGTGCACTCTTTGATGCATTGCTTTAGAAGGGG 93

      42 ySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgA 59
         |:::::::|||   ||||||          |||      :::
      94 AGAGTCTGGGTGTCTTTGCTACTGTTTCTATTTCCGTAAGAAAAATGGCC 143

      59 laCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLys 75
                 |||::::::   :::      :::|||::::::   |||
     144 CATGAAGAAACTTAAAAACTGCAAGTTTGAATTTCTCCAGGCAGGCAAAA 193

      76 ......................TrpLeuSerIle............Al 80
                              ||||||:::|||           ::
     194 ATGAGGTAAATAAAGTAAAGGACTTGGCTGGCAATTCCCAAAGAAGCAGG 243

      80 aTyrGlyArgCysGln...LysIleGlyArgGlnLysLeuAlaLysLysM 96
         :||||||      :::   ::::::||||||:::         :::|||:
     244 CTATGGAGGGACGAGGTCCCGAGTAGGTAGGAAGGCTTATCGAGAGAAGC 293

      96 et 96
         ::
     294 TG 295

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-5019

seq_documentation_block:
; Sequence 5019, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5019
;   LENGTH: 461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(461)
;   OTHER INFORMATION: n = A,T,C or G
US-09-922-340-5019


alignment_scores:
            Quality:   71.50              Length:     84
              Ratio:  1.663                Gaps:      3
 Percent Similarity:  51.190   Percent Identity: 25.000

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-922-340-5019   ..

 Align seg 1/1  to: US-09-922-340-5019  from: 1  to: 461

      26 CysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGl 42
         ||||||   :::::::|||   :::          ||||||   :::::::||
      44 TGCTGTCCACATACCCCAGTGCACTCTTTGATGCATTGCTTTAGAAGGGG 93

      42 ySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgA 59
         |:::::::|||   ||||||          |||      :::
      94 AGAGTCTGGGTGTCTTTGCTACTGTTTCTATTTCCGTAAGAAAAATGGCC 143

      59 laCysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLys 75
                 |||::::::   :::      :::|||::::::   |||
     144 CATGAAGAAACTTAAAAACTGCAAGTTTGAATTTCTCCAGGCAGGCAAAA 193

      76 ......................TrpLeuSerIle............Al 80
                              ||||||:::|||           ::
     194 ATGAGGTAAATAAAGTAAAGGACTTGGCTGGCAATTCCCAAAGAAGCAGG 243

      80 aTyrGlyArgCysGln...LysIleGlyArgGlnLysLeuAlaLysLysM 96
         :||||||      :::   ::::::||||||:::         :::|||:
     244 CTATGGAGGGACGAGGTCCCGAGTAGGTAGGAAGGCTTATCGAGAGAAGC 293

      96 et 96
         ::
     294 TG 295

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-2381

seq_documentation_block:
; Sequence 2381, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 2381
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-2381


alignment_scores:
            Quality:   71.50              Length:     98
              Ratio:  1.521                Gaps:      5
 Percent Similarity:  47.959   Percent Identity: 26.531

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-60-325-448-2381/rev   ..

 Align seg 1/1  to reverse of: US-60-325-448-2381  from: 1  to: 2000

       6 LysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCy 22
         |||  |||:::         |||:::::::  |||   ::::::   ||
    1112 AAAAAGAATGTAGATACAACAAAGAGCAAAAACATCAGCAAGTCTACATG 1063

      22 sGluThrGlyCysCysGlnArgAlaPro.................AspA 33
         |   |||:::|||   ::::::  |||                  |||:
    1062 TACAACTGCCTGCCTATCAAAACTTCCATTATCTACAAGAACAATTGATC 1013
```

```
      33 snCysGluSerHisCysAlaGluLysGlySerGluGlySerLeuCysGln 49
         ::|||:::                                    |||
    1012 GGTGTCGG.................................TGCGCC 999

      50 ThrGlnValPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLe 66
         |||::::::            ||||||   :::         :::   ||
     998 ACATCTTTG...........TACAGACCAGCTGCCATTCCACAGCTCCT 961

      66 uThrCysIleTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyA 83
         |   |||:::   :::||||||||||:::   |||           |||
     960 GCTCTGTGTAGTGGGCAAAAATGAATCAACCCTG...........GGAT 923

      83 rgCys..............GlnLysIleGlyArgGlnLysLeu 92
           |||               ::::::|||.|||||||::::::|||
     922 GTTGCAGAGTACTCGAAGAGAAGTCGATTGGACGGCGTAGATTG 879
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-702

```
seq_documentation_block:
; Sequence 702, Application US/09954456
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 702
;   LENGTH: 10172
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-702


alignment_scores:
            Quality:   71.00             Length:    105
              Ratio:  1.315               Gaps:      6
 Percent Similarity:  51.429   Percent Identity: 21.905

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-954-456-702   ..

 Align seg 1/1  to: US-09-954-456-702  from: 1  to: 10172

       8 AsnLeuLeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluTh 24
         |||   ::::::::::   |||      :::::::::      |||:::::
    2416 AACTGTATTGACATTGATGAATGTTTAGTAAACAGACTGCTTTGTGATAA 2465

      24 rGlyCysCysGlnArgAlaPro.......................... 31
         :|||   |||::::::::::|||
    2466 CGGATTGTGCCGAAACACGCCAGGAAGTTACAGCTGTACGTGCCCACCAG 2515

      32 ......................AspAsnCysGlu......SerHisCys 38
                                 ::::::||||||      ::::::|||
    2516 GGTATGTGTTCAGGACTGAGACAGAGACCTGTGAAGATATAAATGAATGT 2565

      39 AlaGluLysGlySerGluGlySerLeuCysGlnThrGlnValPhePheGl 55
            ::::::          ::::::   |||:::::::::::::      ||
    2566 GAAAGCAACCCATGTGTCAATGGGGCCTGCAGAAACAACCTT......GG 2609

      55 yGlnTyrArgAlaCysProCys.....................LeuArgA 65
         |:::::::::   |||   |||                          :
    2610 ATCTTTCAAT...TGTGAATGTTCGCCCGGCAGCAAACTCAGCTCCACAG 2656

      65 snLeuThrCysIleTyrSer...LysAsnGluLysTrpLeuSerIleAla 80
         ::|||   ||||||   |||   |||:::      ||||||:::|||
    2657 GATTGATCTGTATTGACAGCCTGAAGGGGACCTGTTGGCTCAACATCCAG 2706

      81 TyrGlyArgCysGln 85
            :::||||||:::
    2707 GACAGCCGCTGTGAG 2721
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-189-455

```
seq_documentation_block:
; Sequence 455, Application US/09978189
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi, Avi
;  APPLICANT:  Baker Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan
;  APPLICANT:  Ferrara, Napoleon
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, J. Christopher
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Kuo, Sophia S.
;  APPLICANT:  Napier, Mary A.
;  APPLICANT:  Pan, James;
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Shelton, David L.
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2630P1C7
;  CURRENT APPLICATION NUMBER: US/09/978,189
;  CURRENT FILING DATE:  2001-10-15
;  PRIOR APPLICATION NUMBER: 09/918585
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064249
;  PRIOR FILING DATE: 1997-11-03
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066364
;  PRIOR FILING DATE: 1997-11-21
;  PRIOR APPLICATION NUMBER: 60/077450
;  PRIOR FILING DATE: 1998-03-10
;  PRIOR APPLICATION NUMBER: 60/077632
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077641
```

; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-5-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573

```
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
            Quality:   70.00              Length:     68
              Ratio:  2.059                Gaps:      3
 Percent Similarity:  50.000   Percent Identity: 26.471

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-978-189-455   ..

 Align seg 1/1  to: US-09-978-189-455  from: 1  to: 1518

      16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
         |||:::|||::::::||||||   :::|||   |||
     702 TGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.............. 737

      32 pAsnCysGluSerHis..............CysAlaGluLysGlySerG 44
            |||       |||             |||          :::|
     738 ....TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAG 783

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGln............ 56
         |||||:::::::|||      :::          |||:::
     784 AAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAA 833

      57 ...TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLy 72
           ::::::   |||   |||   ::::::|||:::|||      :::||
     834 ATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA 883

      72 sAsn 73
         |:::
     884 AGAT 887

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-192-455

seq_documentation_block:
; Sequence 455, Application US/09978192
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi, Avi
;  APPLICANT:  Baker Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan
;  APPLICANT:  Ferrara, Napoleon
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, J. Christopher
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Kuo, Sophia S.
;  APPLICANT:  Napier, Mary A.
;  APPLICANT:  Pan, James;
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Shelton, David L.
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2630P1C9
;  CURRENT APPLICATION NUMBER: US/09/978,192
;  CURRENT FILING DATE:  2001-10-15
;  PRIOR APPLICATION NUMBER: 09/918585
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064249
;  PRIOR FILING DATE: 1997-11-03
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066364
;  PRIOR FILING DATE: 1997-11-21
;  PRIOR APPLICATION NUMBER: 60/077450
;  PRIOR FILING DATE: 1998-03-10
;  PRIOR APPLICATION NUMBER: 60/077632
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077641
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077649
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077791
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/078004
;  PRIOR FILING DATE: 1998-03-13
;  PRIOR APPLICATION NUMBER: 60/078886
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078936
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078939
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/079294
;  PRIOR FILING DATE: 1998-03-25
;  PRIOR APPLICATION NUMBER: 60/079656
;  PRIOR FILING DATE: 1998-03-26
;  PRIOR APPLICATION NUMBER: 60/079664
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079663
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079728
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079786
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079920
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/079923
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/080105
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080107
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080165
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080194
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080327
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080328
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080333
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080334
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/081070
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081049
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081071
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081195
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081203
```

```
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081229
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081955
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081817
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081819
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081952
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081838
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/082568
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/082569
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/082704
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082804
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082700
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082797
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/082796
;  PRIOR FILING DATE: 1998-04-23
;  PRIOR APPLICATION NUMBER: 60/083336
;  PRIOR FILING DATE: 1998-04-27
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/083392
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083495
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083496
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083499
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083545
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083554
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083558
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083559
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083500
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/083742
;  PRIOR FILING DATE: 1998-04-30
;  PRIOR APPLICATION NUMBER: 60/084366
;  PRIOR FILING DATE: 1998-05-05
;  PRIOR APPLICATION NUMBER: 60/084414
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: 60/084441
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: 60/084637
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084639
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084640
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084598
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-5-07
;  PRIOR APPLICATION NUMBER: 60/084627
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084643
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/085339
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085338
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085323
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085582
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085700
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085689
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085580
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085573
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
             Quality:   70.00              Length:     68
               Ratio:  2.059                Gaps:      3
 Percent Similarity:  50.000   Percent Identity: 26.471

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-978-192-455   ..

 Align seg 1/1  to: US-09-978-192-455  from: 1  to: 1518

      16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
         |||:::|||::::::|||||   :::|||   |||
     702 TGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.............. 737

      32 pAsnCysGluSerHis..............CysAlaGluLysGlySerG 44
            |||      |||               |||           :::|
     738 ....TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAG 783

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGln............ 56
         |||||::::::|||      :::         |||:::
     784 AAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAA 833

      57 ...TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLy 72
            ::::::   |||   |||   ::::::|||:::|||      :::||
     834 ATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA 883

      72 sAsn 73
         |:::
     884 AGAT 887

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-697-455

 seq_documentation_block:
 ; Sequence 455, Application US/09978697
 ; GENERAL INFORMATION:
 ;  APPLICANT: Ashkenazi, Avi
 ;  APPLICANT:  Baker Kevin P.
 ;  APPLICANT:  Botstein, David
 ;  APPLICANT:  Desnoyers, Luc
 ;  APPLICANT:  Eaton, Dan
 ;  APPLICANT:  Ferrara, Napoleon
 ;  APPLICANT:  Filvaroff, Ellen
 ;  APPLICANT:  Fong, Sherman
 ;  APPLICANT:  Gao, Wei-Qiang
 ;  APPLICANT:  Gerber, Hanspeter
 ;  APPLICANT:  Gerritsen, Mary E.
 ;  APPLICANT:  Goddard, Audrey
 ;  APPLICANT:  Godowski, Paul J.
 ;  APPLICANT:  Grimaldi, J. Christopher
 ;  APPLICANT:  Gurney, Austin L.
 ;  APPLICANT:  Hillan, Kenneth J
 ;  APPLICANT:  Kljavin, Ivar J.
```

; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630P1C27
; CURRENT APPLICATION NUMBER: US/09/978,697
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637

```
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084639
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084640
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084598
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-5-07
;  PRIOR APPLICATION NUMBER: 60/084627
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/084643
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/085339
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085338
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085323
;  PRIOR FILING DATE: 1998-05-13
;  PRIOR APPLICATION NUMBER: 60/085582
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085700
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085689
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085580
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085573
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085704
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/085697
;  PRIOR FILING DATE: 1998-05-15

alignment_scores:
            Quality:   70.00              Length:      68
              Ratio:  2.059                Gaps:       3
 Percent Similarity:  50.000   Percent Identity: 26.471

alignment_block:
 US-09-215-435-179_COPY_24_121 x US-09-978-697-455   ..

 Align seg 1/1  to: US-09-978-697-455  from: 1  to: 1518

      16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
         |||:::|||:::::::|||||   :::|||   |||
     702 TGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.............. 737

      32 pAsnCysGluSerHis..............CysAlaGluLysGlySerG 44
             |||      |||               |||      :::|
     738 ....TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAG 783

      44 luGlySerLeuCysGlnThrGlnValPhePheGlyGln............. 56
         |||||:::::::|||      :::         |||:::
     784 AAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAA 833

      57 ...TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLy 72
           ::::::   |||   |||   ::::::|||:::|||      :::||
     834 ATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA 883

      72 sAsn 73
         |:::
     884 AGAT 887

 seq_name: /cgnl_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-824-455

 seq_documentation_block:
 ; Sequence 455, Application US/09978824
 ; GENERAL INFORMATION:
 ;  APPLICANT: Ashkenazi, Avi
 ;  APPLICANT:  Baker Kevin P.
 ;  APPLICANT:  Botstein, David
 ;  APPLICANT:  Desnoyers, Luc
 ;  APPLICANT:  Eaton, Dan
 ;  APPLICANT:  Ferrara, Napoleon
 ;  APPLICANT:  Filvaroff, Ellen
 ;  APPLICANT:  Fong, Sherman
 ;  APPLICANT:  Gao, Wei-Qiang
 ;  APPLICANT:  Gerber, Hanspeter
 ;  APPLICANT:  Gerritsen, Mary E.
 ;  APPLICANT:  Goddard, Audrey
 ;  APPLICANT:  Godowski, Paul J.
 ;  APPLICANT:  Grimaldi, J. Christopher
 ;  APPLICANT:  Gurney, Austin L.
 ;  APPLICANT:  Hillan, Kenneth J
 ;  APPLICANT:  Kljavin, Ivar J.
 ;  APPLICANT:  Kuo, Sophia S.
 ;  APPLICANT:  Napier, Mary A.
 ;  APPLICANT:  Pan, James;
 ;  APPLICANT:  Paoni, Nicholas F.
 ;  APPLICANT:  Roy, Margaret Ann
 ;  APPLICANT:  Shelton, David L.
 ;  APPLICANT:  Stewart, Timothy A.
 ;  APPLICANT:  Tumas, Daniel
 ;  APPLICANT:  Williams, P. Mickey
 ;  APPLICANT:  Wood, William I.
 ;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ;  TITLE OF INVENTION:  Acids Encoding the Same
 ;  FILE REFERENCE: P2630P1C14
 ;  CURRENT APPLICATION NUMBER: US/09/978,824
 ;  CURRENT FILING DATE:  2001-10-17
 ;  PRIOR APPLICATION NUMBER: 09/918585
 ;  PRIOR FILING DATE: 2001-07-30
 ;  PRIOR APPLICATION NUMBER: 60/062250
 ;  PRIOR FILING DATE: 1997-10-17
 ;  PRIOR APPLICATION NUMBER: 60/064249
 ;  PRIOR FILING DATE: 1997-11-03
 ;  PRIOR APPLICATION NUMBER: 60/065311
 ;  PRIOR FILING DATE: 1997-11-13
 ;  PRIOR APPLICATION NUMBER: 60/066364
 ;  PRIOR FILING DATE: 1997-11-21
 ;  PRIOR APPLICATION NUMBER: 60/077450
 ;  PRIOR FILING DATE: 1998-03-10
 ;  PRIOR APPLICATION NUMBER: 60/077632
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077641
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077649
 ;  PRIOR FILING DATE: 1998-03-11
 ;  PRIOR APPLICATION NUMBER: 60/077791
 ;  PRIOR FILING DATE: 1998-03-12
 ;  PRIOR APPLICATION NUMBER: 60/078004
 ;  PRIOR FILING DATE: 1998-03-13
 ;  PRIOR APPLICATION NUMBER: 60/078886
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078936
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078910
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/078939
 ;  PRIOR FILING DATE: 1998-03-20
 ;  PRIOR APPLICATION NUMBER: 60/079294
 ;  PRIOR FILING DATE: 1998-03-25
 ;  PRIOR APPLICATION NUMBER: 60/079656
 ;  PRIOR FILING DATE: 1998-03-26
 ;  PRIOR APPLICATION NUMBER: 60/079664
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079689
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079663
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079728
 ;  PRIOR FILING DATE: 1998-03-27
 ;  PRIOR APPLICATION NUMBER: 60/079786
```

; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083554
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-5-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15

alignment_scores:
Quality: 70.00 Length: 68
Ratio: 2.059 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 26.471

alignment_block:
US-09-215-435-179_COPY_24_121 x US-09-978-824-455 ..

Align seg 1/1 to: US-09-978-824-455 from: 1 to: 1518

```
     16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
        |||:::|||:::::::||||||   :::|||   |||
    702 TGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.............. 737

     32 pAsnCysGluSerHis..............CysAlaGluLysGlySerG 44
         |||       |||               |||             :::|
    738 ....TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAG 783

     44 luGlySerLeuCysGlnThrGlnValPhePheGlyGln............. 56
        |||||::::::|||      :::          |||:::
    784 AAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAA 833
```

```
     57 ...TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLy 72
        ::::::   |||   |||   ::::::|||:::|||      :::||
    834 ATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA 883

     72 sAsn 73
        |:::
    884 AGAT 887
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-298-88

seq_documentation_block:
; Sequence 88, Application US/09974298
; GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 88
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 2481150CB1
US-09-974-298-88

alignment_scores:
Quality: 70.00 Length: 68
Ratio: 2.059 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 26.471

alignment_block:
US-09-215-435-179_COPY_24_121 x US-09-974-298-88 ..

Align seg 1/1 to: US-09-974-298-88 from: 1 to: 1797

```
     16 CysIleArgAsnGlnAspCysGluThrGlyCysCysGlnArgAlaProAs 32
        |||:::|||:::::::||||||   :::|||   |||
    716 TGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGT.............. 751

     32 pAsnCysGluSerHis..............CysAlaGluLysGlySerG 44
           |||      |||              |||            :::|
    752 ....TGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAG 797

     44 luGlySerLeuCysGlnThrGlnValPhePheGlyGln..... 56
        |||||:::::::|||      :::         |||:::
    798 AAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAA 847

     57 ...TyrArgAlaCysProCysLeuArgAsnLeuThrCysIleTyrSerLy 72
        ::::::   |||   |||   :::::::|||:::|||      :::||
    848 ATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA 897

     72 sAsn 73
        |:::
    898 AGAT 901
```

```
OM of: US-09-215-435-179_COPY_24_121 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:30 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-179_COPY_24_121
Query length: 98
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est15:AI018488      -  480.00  959.38  2.6e-44   491  ! AI018488 ov63f08.s1 Soares_test
gb_gss29:AZ575911      +  196.00  399.53  3.9e-13   127  ! AZ575911 AST-T22D0047 Genetrap
gb_est88:BF509739      -  164.50  327.17  4.2e-09   293  ! BF509739 UI-H-BI4-apg-b-05-0-UI
gb_est69:BE047055      -  137.00  265.23  1.2e-05   540  ! BE047055 hq55g09.x1 NCI_CGAP_Pa
gb_est19:AI366744      -  135.00  262.33  1.7e-05   484  ! AI366744 qq41a03.x1 Soares_NhHM
gb_est88:BF507781      -  135.00  262.01  1.8e-05   499  ! BF507781 UI-H-BI4-apv-b-04-0-UI
gb_est81:BE969660      +  135.00  261.42  1.9e-05   528  ! BE969660 601679794F1 NIH_MGC_78
gb_est69:BE047063      -  135.00  261.40  1.9e-05   529  ! BE047063 hq55h07.x1 NCI_CGAP_Pa
gb_est81:BE969913      +  135.00  261.33  1.9e-05   533  ! BE969913 601680001F1 NIH_MGC_78
gb_est81:BE969784      +  135.00  261.13  2.0e-05   543  ! BE969784 601679544F1 NIH_MGC_78
gb_est81:BE969920      +  135.00  260.50  2.2e-05   577  ! BE969920 601680009F1 NIH_MGC_78
gb_est81:BE969726      +  135.00  258.76  2.7e-05   682  ! BE969726 601679676F1 NIH_MGC_78
gb_est81:BE969942      +  135.00  258.21  2.9e-05   719  ! BE969942 601680032F1 NIH_MGC_78
gb_est81:BE970098      +  135.00  258.01  3.0e-05   733  ! BE970098 601679841F1 NIH_MGC_78
gb_est81:BE970031      +  135.00  257.74  3.1e-05   752  ! BE970031 601680144F1 NIH_MGC_78
gb_est81:BE970121      +  135.00  257.58  3.1e-05   764  ! BE970121 601679866F1 NIH_MGC_78
gb_est81:BE970243      +  135.00  257.47  3.2e-05   772  ! BE970243 601679987F1 NIH_MGC_78
gb_est49:AW583164      +  131.00  255.92  3.9e-05   412  ! AW583164 ia11f06.y1 Human Pancr
gb_est81:BE970103      +  127.00  241.72  0.0002    742  ! BE970103 601679846F1 NIH_MGC_78
gb_est19:AI411581      -  125.00  245.41  0.0001    353  ! AI411581 EST239875 Normalized r
gb_est49:AW584008      -  125.00  241.39  0.0003    519  ! AW584008 ia08f03.x1 Human Pancr
gb_est68:BB565103      +  124.00  245.47  0.0001    289  ! BB565103 BB565103 RIKEN full-le
gb_est53:AW966719      +  124.00  240.21  0.0003    479  ! AW966719 EST378793 MAGE reseque
gb_est81:BE970191      +  124.00  238.85  0.0003    546  ! BE970191 601679930F1 NIH_MGC_78
gb_est99:BG311607      +  123.00  237.76  0.0004    499  ! BG311607 ib26f02.y1 Melton Mous
gb_est49:AW583194      +  123.00  237.31  0.0004    521  ! AW583194 ia12c01.y1 Human Pancr
gb_est81:BE970198      +  123.00  236.64  0.0005    556  ! BE970198 601679939F1 NIH_MGC_78
gb_est53:AW915561      +  122.00  239.34  0.0003    353  ! AW915561 EST346865 Normalized r
gb_est12:AA851206      -  122.00  237.84  0.0004    408  ! AA851206 EST193974 Normalized r
gb_est17:AI178610      -  122.00  237.49  0.0004    422  ! AI178610 EST222289 Normalized r
gb_est98:BG271031      +  122.00  235.99  0.0005    487  ! BG271031 ib03c06.y1 Melton Mous
gb_est49:AW582939      +  122.00  235.41  0.0005    515  ! AW582939 ia08b04.y1 Human Pancr
gb_est19:AI385475      +  121.00  235.46  0.0005    422  ! AI385475 mj69b11.y1 Soares mous
gb_est98:BG270511      +  121.00  235.27  0.0006    430  ! BG270511 ia85b12.y1 Melton Mous
gb_est97:BG141967      -  121.00  235.22  0.0006    432  ! BG141967 ia85b12.x1 Melton Mous
gb_est98:BG270629      +  121.00  235.17  0.0006    434  ! BG270629 ia86f04.y1 Melton Mous
gb_est18:AI326929      -  121.00  234.98  0.0006    442  ! AI326929 mj69b11.x1 Soares mous
gb_est97:BG141777      -  121.00  234.96  0.0006    443  ! BG141777 ia94b06.x1 Melton Mous
gb_est99:BG311902      -  121.00  234.75  0.0006    452  ! BG311902 ia98a11.x1 Melton Mous
gb_est99:BG312118      -  121.00  234.15  0.0006    479  ! BG312118 ib03c06.x1 Melton Mous
gb_est97:BG143141      +  121.00  234.08  0.0006    482  ! BG143141 ia94b06.y1 Melton Mous
gb_est99:BG311196      +  121.00  234.02  0.0006    485  ! BG311196 ib21c03.y1 Melton M
gb_est99:BG311655      +  121.00  234.02  0.0006    485  ! BG311655 ib27d10.y1 Melton M
gb_est100:BG412963     +  121.00  233.99  0.0006    486  ! BG412963 ia82e06.y1 Melton M
gb_est98:BG270331      +  121.00  233.95  0.0007    488  ! BG270331 ib13e11.y1 Melton M

seq_name: gb_est15:AI018488

seq_documentation_block:
LOCUS       AI018488      491 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov63f08.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642023
            3', mRNA sequence.
ACCESSION   AI018488
VERSION     AI018488.1  GI:3233007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 630   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1642023"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      129 a    115 c    133 g    113 t      1 others
ORIGIN


alignment_scores:
            Quality:  480.00             Length:     97
              Ratio:  5.161                Gaps:      1
 Percent Similarity:  95.876   Percent Identity: 94.845

alignment_block:
 US-09-215-435-179_COPY_24_121 x AI018488/rev   ..

 Align seg 1/1  to reverse of: AI018488  from: 1  to: 491

       3 SerProThrLysTyrAsnLeuLeuGluLeuLysGluSerCysIleArgAs 19
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     442 TCTCCAACAAAATACAACCTTTTGGAGCTCAAGGAGTCTTGCATCCGGAA 393

      19 nGlnAspCysGluThrGlyCysCysGlnArgAlaProAspAsnCysGluS 36
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    392 CCAGGACTGCGAGACTGGCTGCTGCCAACGTGCTCCAGACAATTGCGAGT 343

     36 erHis.CysAlaGluLysGlySerGluGlySerLeuCysGlnThrGlnVa 52
        |||||    ||||||||||||||||||||||||||||||||||||||||
    342 CGCACTNGCGCGGAGAAGGGGTCCGAGGGCAGTCTGTGTCAAACGCAGGT 293

     52 lPhePheGlyGlnTyrArgAlaCysProCysLeuArgAsnLeuThrCysI 69
        |||||||||||||||||||||||||||||||   ||||||||||||||||
    292 GTTCTTTGGCCAATATAGAGCGTGTCCCTGCTG.CGGAACCTGACTTGTA 244

     69 leTyrSerLysAsnGluLysTrpLeuSerIleAlaTyrGlyArgCysGln 85
        |||||||||||||||||||||||||||||:::   ||||||||||||||||
    243 TATATTCAAAGAATGAGAAATGGCTTAGCATGCCCTATGGCCGTTGTCAG 194

     86 LysIleGlyArgGlnLysLeuAlaLysLysMetPhePhe 98
        |||||||||||||||||||||||||||||||||||||||
    193 AAAATTGGAAGGCAGAAGTTGGCTAAGAAAATGTTCTTC 155

seq_name: gb_gss29:AZ575911

seq_documentation_block:
LOCUS       AZ575911      127 bp    DNA             GSS       06-DEC-2000
DEFINITION  AST-T22D0047 Genetrap T47D Human Breast Carcinoma Library Homo
            sapiens genomic 5', DNA sequence.
ACCESSION   AZ575911
VERSION     AZ575911.1  GI:11562222
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127)
  AUTHORS   Henkel,G., Liyanage,M., Pratt,E., Huang,D., Riley,M., Bernardino,A.
            , Durick,K. and Pollok,B.
  TITLE     Exon-trap tags from a T47D GenomeScreen(TM) Library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Greg Henkel
            Gene Expression
            Aurora Biosciences Corp.
            11010 Torreyana Road, San Diego, CA 92121, USA
            Tel: 8584048436
            Fax: 8584046719
            Email: henkelg@aurorabio.com
            Pools of cells were isolated from a GenomeScreen(TM) library. The
            library of cells was generated by retroviral integration of a gene
            tagging element consisting of: 1) A promoterless beta-lactamase
            proceeded by a splice acceptor as a reporter for gene expression;
            2) A promoter driving neomycin resistance followed by a splice
            donor to trap downstream exons. 3' RACE from neomycin gene was
            performed using total RNA from isolated pools. Output was shotgun
            cloned in pAmp-1 and used to transform DH5-alpha competent
            bacteria. 5' ends of reported sequences were immediately preceded
            by splice donor from the trapping construct.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .127
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Genetrap T47D Human Breast Carcinoma Library"
                     /tissue_type="Carcinoma"
                     /cell_type="Epithelial"
                     /cell_line="T47D"
                     /note="Organ: Breast; Vector: pAmp-1; 3' RACE of total RNA
                     from genetrap pools; shotgun clone in pAmp-1 and  used to
                     transform DH5-alpha competent bacteria."
BASE COUNT       28 a     34 c     43 g     21 t      1 others
ORIGIN


alignment_scores:
            Quality:  196.00          Length:     40
              Ratio:  5.026             Gaps:      0
 Percent Similarity:  97.500   Percent Identity: 97.500
```

```
alignment_block:
US-09-215-435-179_COPY_24_121 x AZ575911   ..

Align seg 1/1  to: AZ575911  from: 1  to: 127

     11 GluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCysCy 27
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 GAGCTCAAGGAGTCTTGCATCCGGAACCAGGACTGCGAGACTGGCTGCTG 50

     27 sGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlySerG 44
        ||||||||||||||||   ||||||||||||||||||||||||||||||
     51 CCAACGTGCTCCAGACAA.TGCGAGTCGCACTGCGCGGAGAAGGGGTCCN 99

     44 luGlySerLeuCysGlnThr 50
        |||||||||||||||||
    100 AGGGCAGTCTGTGTCAAACG 119

seq_name: gb_est88:BF509739

seq_documentation_block:
LOCUS       BF509739      293 bp    mRNA            EST       06-DEC-2000
DEFINITION  UI-H-BI4-apg-b-05-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3087296 3', mRNA sequence.
ACCESSION   BF509739
VERSION     BF509739.1  GI:11593037
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3087296"
                     /clone_lib="NCI_CGAP_Sub8"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive  hybridization with a driver
                     comprising: a pool of clones from  NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2732833-2737415, 3068040-3069191; 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729326; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728969-2733190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238,  3081864-3084550
                     ; 25% of the driver population).  Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB=NCI_CGAP_Br2
```

```
                    TAG_TISSUE=breast
                    TAG_SEQ=AAACC"
BASE COUNT       63 a     75 c     72 g     83 t
ORIGIN


alignment_scores:
            Quality:  164.50              Length:     38
              Ratio:  5.306                 Gaps:      1
 Percent Similarity:  81.579   Percent Identity: 78.947

alignment_block:
 US-09-215-435-179_COPY_24_121 x BF509739/rev   ..

 Align seg 1/1  to reverse of: BF509739  from: 1  to: 293

      25 GlyCysCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLy 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 GGCTGCTGCCAACGTGCTCCAGACAATTGCGAGTCGCACTGCGCGGAGAA 235

      41 sGlySerGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrA 58
         ||||||||||||||||||||||||||||||:::
     234 GGGGTCCGAGGGCAGTCTGTGTCAAACGCAGCTG..............C 200

      58 rgAlaCysProCys 62
            |||   ||||||
     199 CAGCTACTCCATGC 186

seq_name: gb_est69:BE047055

seq_documentation_block:
LOCUS       BE047055      540 bp    mRNA            EST       08-JUN-2000
DEFINITION  hq55g09.x1 NCI_CGAP_Pan3 Homo sapiens cDNA clone IMAGE:3123328 3'
            similar to gb:J02883 COLIPASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   BE047055
VERSION     BE047055.1  GI:8364108
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3123328"
                     /clone_lib="NCI_CGAP_Pan3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      123 a    131 c    161 g    124 t      1 others
ORIGIN
```

```
alignment_scores:
            Quality:  137.00              Length:     87
              Ratio:  2.362                 Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE047055/rev   ..

 Align seg 1/1  to reverse of: BE047055  from: 1  to: 540

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::||
     437 TTGGAGACCGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 388

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                :::::::|||::::::   :::|
     387 CTGCCAGCATTCAAGTGCGTTGGGCTTGGCCCGCTGCACATCCATGGCCA 338

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::::|||   |||
     337 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTTTATGGGATTTACTACAAG 288

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         ::::::::::|||
     287 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 247

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
         :::              :::      :::|||   |||:::     |
     246 CATTGTGGGCTCCATCACCAACACCAACTTTGGCATTTGCCATGACGNTG 197

      88 lyArgGlnLys 91
         |||||:::|||
     196 GACGCTCCAAG 186

seq_name: gb_est19:AI366744

seq_documentation_block:
LOCUS       AI366744      484 bp    mRNA            EST       13-FEB-1999
DEFINITION  qq41a03.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1935052
            3' similar to gb:J02883 COLIPASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AI366744
VERSION     AI366744.1  GI:4136489
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 715   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1935052"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
```

were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 105 a 121 c 159 g 99 t
ORIGIN

alignment_scores:
Quality: 135.00 Length: 87
Ratio: 2.328 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 32.184

alignment_block:
US-09-215-435-179_COPY_24_121 x AI366744/rev ..

Align seg 1/1 to reverse of: AI366744 from: 1 to: 484

```
     10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
        ||||||      |||   |||:::::::::   :::|||:::::::::||
    423 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 374

     26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
        |||||||::::::                 :::::::|||:::::::   :::|
    373 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 324

     43*erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
        |||||:::|||   |||:::::::::::::::::::|||   |||
    323 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 274

     60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
        |||||||||   |||:::|||||||||         ::::::::::|||
    273 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 233

     76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
         :::               :::      :::|||   |||:::      |
    232 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 183

     88 lyArgGlnLys 91
        |||||:::|||
    182 GACGCTCCAAG 172
```

seq_name: gb_est88:BF507781

seq_documentation_block:
LOCUS BF507781 499 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-BI4-apv-b-04-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone IMAGE:3088471 3', mRNA sequence.
ACCESSION BF507781
VERSION BF507781.1 GI:11591079
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 499)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3088471"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8 is a subtracted library derived from NCI_CGAP_Sub5. The NCI_CGAP_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE clone Ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone Ids 2723592-2729326; 25% of the driver population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids 2728969-2733190; 25% of the driver population), and NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550 ; 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_LIB=NCI_CGAP_Brn26
TAG_TISSUE=brain
TAG_SEQ=ATAGG"

BASE COUNT 106 a 121 c 156 g 115 t 1 others
ORIGIN

alignment_scores:
Quality: 135.00 Length: 87
Ratio: 2.328 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 32.184

alignment_block:
US-09-215-435-179_COPY_24_121 x BF507781/rev ..

Align seg 1/1 to reverse of: BF507781 from: 1 to: 499

```
     10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
        ||||||      |||   |||:::::::::   :::|||:::::::::||
    439 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 390

     26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
        |||||||::::::                 :::::::|||:::::::   :::|
    389 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 340

     43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
        |||||:::|||   |||:::::::::::::::::::|||   |||
    339 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 290

     60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
        |||||||||   |||:::|||||||||         ::::::::::|||
    289 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 249

     76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
         :::               :::      :::|||   |||:::      |
    248 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 199

     88 lyArgGlnLys 91
        |||||:::|||
    198 GACGCTCCAAG 188
```

seq_name: gb_est81:BE969660

seq_documentation_block:
LOCUS BE969660 528 bp mRNA EST 04-OCT-2000
DEFINITION 601679794F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949990 5', mRNA sequence.
ACCESSION BE969660

```
VERSION     BE969660.1  GI:10582593
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM815  row: 1  column: 23
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949990"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      117 a    168 c    133 g    110 t
ORIGIN


alignment_scores:
            Quality:  135.00            Length:     87
              Ratio:  2.328               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE969660   ..

 Align seg 1/1  to: BE969660  from: 1  to: 528

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| |
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::              ::::::|||::::::   :::|
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::|||   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         :::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::             :::      :::|||   |||:::      |
     289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

      88 lyArgGlnLys 91
```

```
         |||||:::|||
     339 GACGCTCCAAG 349

seq_name: gb_est69:BE047063

seq_documentation_block:
LOCUS       BE047063      529 bp    mRNA            EST       08-JUN-2000
DEFINITION  hq55h07.x1 NCI_CGAP_Pan3 Homo sapiens cDNA clone IMAGE:3123325 3'
            similar to gb:J02883 COLIPASE PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   BE047063
VERSION     BE047063.1  GI:8364116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3123325"
                     /clone_lib="NCI_CGAP_Pan3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      117 a    128 c    167 g    117 t
ORIGIN


alignment_scores:
            Quality:  135.00            Length:     87
              Ratio:  2.328               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE047063/rev   ..

 Align seg 1/1  to reverse of: BE047063  from: 1  to: 529

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| |
     427 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 378

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::              ::::::|||::::::   :::|
     377 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 328

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::|||   |||
     327 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 278

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         :::::::::|||
```

```
    277 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 237

     76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::               :::      :::|||   |||:::      |
    236 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 187

     88 lyArgGlnLys 91
        |||||:::|||
    186 GACGCTCCAAG 176

seq_name: gb_est81:BE969913

seq_documentation_block:
LOCUS       BE969913      533 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680001F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950113 5',
            mRNA sequence.
ACCESSION   BE969913
VERSION     BE969913.1  GI:10582846
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: b  column: 02
            High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950113"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      118 a    169 c    132 g    114 t
ORIGIN


alignment_scores:
            Quality:  135.00             Length:     87
              Ratio:  2.328                Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE969913   ..

 Align seg 1/1  to: BE969913  from: 1  to: 533

     10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
        ||||||      |||   |||:::::::::   :::|||:::::::::||
     98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

     26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
        |||||||::::::              :::::::|||::::::   :::|
    148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

     43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
        |||||:::|||   |||:::::::::::::::::::|||   |||
    198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

     60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
        |||||||||   |||:::|||||||||         ::::::::::|||
    248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

     76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::               :::      :::|||   |||:::      |
    289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

     88 lyArgGlnLys 91
        |||||:::|||
    339 GACGCTCCAAG 349

seq_name: gb_est81:BE969784

seq_documentation_block:
LOCUS       BE969784      543 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679544F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949859 5',
            mRNA sequence.
ACCESSION   BE969784
VERSION     BE969784.1  GI:10582717
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM815  row: g  column: 12
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949859"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      118 a    176 c    133 g    115 t      1 others
ORIGIN


alignment_scores:
            Quality:  135.00             Length:     87
              Ratio:  2.328                Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184
```

```
alignment_block:
 US-09-215-435-179_COPY_24_121 x BE969784   ..

 Align seg 1/1  to: BE969784  from: 1  to: 543

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||::::::::::||
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||:::::               ::::::|||::::::   :::|
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||::::::::::::::::::|||   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::||||||||         :::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::               :::      :::|||   |||:::      |
     289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

      88 lyArgGlnLys 91
         |||||:::|||
     339 GACGCTCCAAG 349

seq_name: gb_est81:BE969920

seq_documentation_block:
LOCUS       BE969920      577 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680009F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950115 5',
            mRNA sequence.
ACCESSION   BE969920
VERSION     BE969920.1  GI:10582853
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: b  column: 04
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950115"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      139 a    174 c    144 g    120 t
ORIGIN


alignment_scores:
            Quality:  135.00             Length:     87
              Ratio:  2.328                Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE969920   ..

 Align seg 1/1  to: BE969920  from: 1  to: 577

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||::::::::::||
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||:::::               ::::::|||::::::   :::|
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||::::::::::::::::::|||   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::||||||||         :::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::               :::      :::|||   |||:::      |
     289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

      88 lyArgGlnLys 91
         |||||:::|||
     339 GACGCTCCAAG 349

seq_name: gb_est81:BE969726

seq_documentation_block:
LOCUS       BE969726      682 bp    mRNA            EST       04-OCT-2000
DEFINITION  601679676F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949891 5',
            mRNA sequence.
ACCESSION   BE969726
VERSION     BE969726.1  GI:10582659
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM815  row: h  column: 20
            High quality sequence stop: 547.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949891"
```

```
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      173 a    198 c    191 g    120 t
ORIGIN


alignment_scores:
            Quality:  135.00              Length:     87
              Ratio:  2.328                 Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE969726   ..

 Align seg 1/1  to: BE969726  from: 1  to: 682

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::::   :::|||:::::::::||
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                :::::::|||::::::   :::|
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::::::   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         ::::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::             :::      :::|||   |||:::      |
     289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

      88 lyArgGlnLys 91
         |||||:::|||
     339 GACGCTCCAAG 349

 seq_name: gb_est81:BE969942

 seq_documentation_block:
 LOCUS       BE969942      719 bp    mRNA            EST       04-OCT-2000
 DEFINITION  601680032F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950455 5',
             mRNA sequence.
 ACCESSION   BE969942
 VERSION     BE969942.1  GI:10582875
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 719)
   AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
   TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
   JOURNAL   Unpublished (1999)
 COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONETECH Laboratories, Inc.
              cDNA Library Preparation: CLONETECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM816  row: p  column: 08
             High quality sequence stop: 524.
 FEATURES             Location/Qualifiers
      source          1. .719
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3950455"
                      /clone_lib="NIH_MGC_78"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                      Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                      ); 5' and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C, G, or T). Average
                      insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies
                      contained inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
 BASE COUNT      160 a    233 c    191 g    135 t
 ORIGIN


 alignment_scores:
             Quality:  135.00              Length:     87
               Ratio:  2.328                 Gaps:      2
  Percent Similarity:  66.667   Percent Identity: 32.184

 alignment_block:
  US-09-215-435-179_COPY_24_121 x BE969942   ..

  Align seg 1/1  to: BE969942  from: 1  to: 719

       10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
          ||||||      |||   |||:::::::::   :::|||:::::::::||
       98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

       26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
          |||||||::::::                :::::::|||::::::   :::|
      148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

       43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
          |||||:::|||   |||:::::::::::::::::::::   |||
      198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

       60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
          |||||||||   |||:::|||||||||         ::::::::::|||
      248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

       76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
           :::             :::      :::|||   |||:::      |
      289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

       88 lyArgGlnLys 91
          |||||:::|||
      339 GACGCTCCAAG 349

  seq_name: gb_est81:BE970098

  seq_documentation_block:
  LOCUS       BE970098      733 bp    mRNA            EST       04-OCT-2000
  DEFINITION  601679841F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950098 5',
              mRNA sequence.
  ACCESSION   BE970098
  VERSION     BE970098.1  GI:10583131
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: a  column: 11
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950098"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      171 a    218 c    210 g    134 t
ORIGIN


alignment_scores:
            Quality:  135.00            Length:     87
              Ratio:  2.328               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE970098   ..

 Align seg 1/1  to: BE970098  from: 1  to: 733

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| 
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                ::::::|||::::::   :::| 
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::|||   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         ::::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
          :::               :::      :::|||   |||:::      |
     289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

      88 lyArgGlnLys 91
         |||||:::|||
     339 GACGCTCCAAG 349

seq_name: gb_est81:BE970031

seq_documentation_block:
```

```
LOCUS       BE970031      752 bp    mRNA            EST       04-OCT-2000
DEFINITION  601680144F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3950266 5',
            mRNA sequence.
ACCESSION   BE970031
VERSION     BE970031.1  GI:10582964
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM816  row: h  column: 11
            High quality sequence stop: 523.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3950266"
                     /clone_lib="NIH_MGC_78"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      168 a    255 c    201 g    128 t
ORIGIN


alignment_scores:
            Quality:  135.00            Length:     87
              Ratio:  2.328               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 32.184

alignment_block:
 US-09-215-435-179_COPY_24_121 x BE970031   ..

 Align seg 1/1  to: BE970031  from: 1  to: 752

      10 LeuGluLeuLysGluSerCysIleArgAsnGlnAspCysGluThrGlyCy 26
         ||||||      |||   |||:::::::::   :::|||:::::::::|| 
      98 CTGGAGAACGGTGAGCTCTGCATGAATAGTGCCCAGTGTAAGAGCAATTG 147

      26 sCysGlnArgAlaProAspAsnCysGluSerHisCysAlaGluLysGlyS 43
         |||||||::::::                ::::::|||::::::   :::| 
     148 CTGCCAGCATTCAAGTGCGCTGGGCCTGGCCCGCTGCACATCCATGGCCA 197

      43 erGluGlySerLeuCysGlnThrGlnValPhePheGlyGlnTyrArgAla 59
         |||||:::|||   |||:::::::::::::::::|||   |||
     198 GCGAGAACAGCGAGTGCTCTGTCAAGACGCTCTATGGGATTTACTACAAG 247

      60 CysProCysLeuArgAsnLeuThrCysIleTyrSerLysAsnGluLysTr 76
         |||||||||   |||:::|||||||||         ::::::::::|||
     248 TGTCCCTGTGAGCGTGGCCTGACCTGT.........GAGGGAGACAAGAC 288

      76 pLeu..............SerIleAlaTyrGlyArgCysGlnLysIleG 88
```

```
     :::                 :::       :::|||   |||:::      |
289 CATCGTGGGCTCCATCACCAACACCAACTTTGGCATCTGCCATGACGCTG 338

 88 lyArgGlnLys 91
    |||||:::|||
339 GACGCTCCAAG 349
```

```
OM of: US-09-215-435-225 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:48 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat2:AX083412       +  1224.00 2255.32  2.2e-117  681  ! AX083412 Sequence 104 from Pate
gb_pat2:AX083422       +  1224.00 2252.82  3.0e-117  897  ! AX083422 Sequence 114 from Pate
gb_pat1:AX060293       +  1220.00 2245.07  8.2e-117  932  ! AX060293 Sequence 1 from Patent
gb_sts2:G27363         -   444.00  816.59  3.0e-37   352  ! G27363 human STS SHGC-31033, se
gb_htg13:AC055854      -   262.50  424.04  2.2e-15   179814  ! AC055854 Homo sapiens chromo
gb_htg12:AC037441      -   246.00  394.38  9.9e-14   162795  ! AC037441 Homo sapiens clone
gb_htg17:AC084081      -   246.00  394.01  1.0e-13   169645  ! AC084081 Homo sapiens chromo
gb_ro2:RNPBP           +   237.50  424.35  2.1e-15   1037  ! X71873 R.norvegicus mRNA for p
gb_pat2:E05646         +   237.50  424.26  2.1e-15   1047  ! E05646 cDNA encoding rat hippo
gb_ro2:RNPEABP         +   237.50  424.02  2.2e-15   1075  ! X75253 R.norvegicus phosphatid
gb_in4:OVD1PROTN       +   234.50  423.57  2.3e-15   611  ! X87991 O.volvulus mRNA for d1 p
gb_in3:CELF40A3        +   230.00  377.18  9.0e-13   41155  ! AF016423 Caenorhabditis elega
gb_htg1:AC006915       +   230.00  366.50  3.5e-12   134095  ! AC006915 Caenorhabditis eleg
gb_htg4:AC012978       -   229.00  371.40  1.9e-12   63551  ! AC012978 Drosophila melanogas
gb_in1:AC005891        +   229.00  364.46  4.6e-12   136933  ! AC005891 Drosophila melanoga
gb_in1:AE003637        +   229.00  360.02  8.1e-12   223646  ! AE003637 Drosophila melanoga
gb_ro1:AF226629        +   227.50  404.47  2.7e-14   1205  ! AF226629 Rattus norvegicus pho
gb_ro1:AF300422        +   225.00  406.70  2.0e-14   564  ! AF300422 Mus musculus phosphati
gb_ro1:AB046417        +   225.00  399.86  4.9e-14   1202  ! AB046417 Mus musculus HCNPpp m
gb_htg6:AC020316       -   222.50  356.99  1.2e-11   82631  ! AC020316 Drosophila melanogas
gb_in2:AE003675        +   222.50  351.85  2.3e-11   145786  ! AE003675 Drosophila melanoga
gb_in4:OVD2PROTN       +   219.00  396.43  7.6e-14   514  ! X87990 O.volvulus mRNA for d1 p
gb_ro1:MMU43206        +   219.00  388.92  2.0e-13   1180  ! U43206 Mus musculus phosphatid
gb_pr10:HUMHRPBP       +   208.50  367.71  3.0e-12   1434  ! D16111 Human mRNA for human ho
gb_pr9:HSPEABP         +   208.50  367.64  3.1e-12   1444  ! X75252 H.sapiens phosphatidyle
gb_pr10:MFPBP          +   208.50  367.64  3.1e-12   1444  ! X73137 M.fascicularis mRNA for
gb_pat2:E05647         +   208.50  367.62  3.1e-12   1447  ! E05647 cDNA encoding human hip
em_pat:E09006          +   207.50  369.10  2.5e-12   1001  ! E09006 Human cDNA encoding a p
gb_htg5:AC016445       +   207.00  327.25  5.4e-10   92569  ! AC016445 Drosophila melanogas
gb_pl2:AF332406        +   206.50  373.14  1.5e-12   522  ! AF332406 Arabidopsis thaliana p
gb_pl2:AF147721        +   206.50  368.96  2.6e-12   829  ! AF147721 Arabidopsis thaliana E
em_pat:E09005          +   205.00  369.66  2.4e-12   564  ! E09005 Human cDNA encoding a ph
gb_pr9:HSRNAPBP        +   205.00  369.56  2.4e-12   570  ! X85033 H.sapiens mRNA for phosp
gb_pr10:S76773         +   202.00  361.86  6.4e-12   723  ! S76773 neuropolypeptide h3 [hum
gb_pat2:I11816         +   201.50  359.77  8.4e-12   822  ! I11816 Sequence 1 from Patent U
gb_in4:ONGANTIG        +   201.50  359.77  8.4e-12   822  ! M27807 O.volvulus O.V.-16 antig
gb_in1:AB035321        +   196.00  353.48  1.9e-11   534  ! AB035321 Dirofilaria immitis mR
gb_htg13:AC062037      -   194.00  296.80  2.7e-08   187355  ! AC062037 Homo sapiens chromo
gb_htg12:AC037441      +   191.50  293.44  4.2e-08   162795  ! AC037441 Homo sapiens clone
gb_htg5:AC016445       -   190.50  296.69  2.7e-08   92569  ! AC016445 Drosophila melanogas
gb_pl2:AF152907        +   182.00  327.66  5.2e-10   528  ! AF152907 Arabidopsis thaliana t
gb_pl1:AB027506        +   182.00  323.91  8.3e-10   799  ! AB027506 Arabidopsis thalian
gb_htg4:AC013131       +   182.00  291.03  5.7e-08   30334  ! AC013131 Drosophila melano
gb_in1:AC004439        +   182.00  281.82  1.8e-07   84055  ! AC004439 Drosophila melano
gb_in1:AE003586        +   182.00  270.13  8.3e-07   306352  ! AE003586 Drosophila melan

seq_name: gb_pat2:AX083412

seq_documentation_block:
LOCUS       AX083412      681 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 104 from Patent WO0112660.
ACCESSION   AX083412
VERSION     AX083412.1  GI:13185249
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 104 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      179 a    186 c    183 g    133 t
ORIGIN


alignment_scores:
           Quality: 1224.00             Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AX083412   ..

 Align seg 1/1  to: AX083412  from: 1  to: 681

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 50

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 100

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 150

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 200

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 250

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 300

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 350

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 400
```

```
    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 450

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 500

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 550

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    551 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 600

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    601 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 650

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    651 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 681

seq_name: gb_pat2:AX083422

seq_documentation_block:
LOCUS       AX083422      897 bp    DNA              PAT       28-FEB-2001
DEFINITION  Sequence 114 from Patent WO0112660.
ACCESSION   AX083422
VERSION     AX083422.1  GI:13185262
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 114 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             99. .782
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC33305.1"
                     /db_xref="GI:13185263"
                     /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
                     GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                     AEPRQRFWRHWLVTDIKGADLKKGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                     VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                     KNQAEIAAC"
BASE COUNT      226 a    253 c    234 g    184 t
ORIGIN


alignment_scores:
            Quality: 1224.00            Length:    227
              Ratio:  5.392              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AX083422   ..

 Align seg 1/1  to: AX083422  from: 1  to: 897

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     99 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 148

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    149 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 198

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    199 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 248

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    249 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 298

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    299 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 348

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    349 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 398

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    399 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 448

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    449 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 498

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    499 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 548

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    549 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 598

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    599 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 648

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    649 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 698

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    699 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 748

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    749 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 779

seq_name: gb_pat1:AX060293

seq_documentation_block:
LOCUS       AX060293      932 bp    DNA              PAT       22-JAN-2001
DEFINITION  Sequence 1 from Patent WO0078802.
ACCESSION   AX060293
VERSION     AX060293.1  GI:12405782
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and
            Herrmann,J.L.
  TITLE     Secreted polypeptides and corresponding polynucleotides
  JOURNAL   Patent: WO 0078802-A 1 28-DEC-2000;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .932
```

```
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
     CDS        113. .796
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC24866.1"
                /db_xref="GI:12405783"
                /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
                GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                AEPRQRFWRHWLVTDIKGADLKEGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                KNQAEIAAC"
BASE COUNT      244 a    257 c    244 g    187 t
ORIGIN


alignment_scores:
            Quality: 1220.00             Length:    227
              Ratio:  5.374               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x AX060293   ..

 Align seg 1/1  to: AX060293  from: 1  to: 932

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     113 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 162

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     163 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 212

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     213 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 262

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     263 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 312

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     313 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 362

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     363 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 412

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     413 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 462

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||:::||||||||||||||||||||||||||
     463 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 512

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 562

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     563 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 612

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     613 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 662

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     663 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 712

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     713 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 762

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     763 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 793

seq_name: gb_sts2:G27363

seq_documentation_block:
LOCUS       G27363         352 bp    DNA             STS       28-JUN-1996
DEFINITION  human STS SHGC-31033, sequence tagged site.
ACCESSION   G27363
VERSION     G27363.1  GI:1396086
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: TTGGAAAAGAAGGGGTTCTG
            Primer B: AAGCACAAAAACCAGGCG
            STS size: 133
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from H83927
            -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="8"
     STS             4. .136
     primer_bind     4. .23
     primer_bind     complement(119. .136)
BASE COUNT       69 a     76 c    110 g     96 t      1 others
ORIGIN


alignment_scores:
            Quality:  444.00             Length:     83
```

```
            Ratio:  5.349                Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-215-435-225 x G27363/rev   ..

Align seq 1/1  to reverse of: G27363  from: 1  to: 352

     145 SerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysVa 161
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     352 AGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCTTCAGGAAGGAAAAGT 303

     161 lIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetA 178
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CATCTCTCTCCTTCCCAAGGAAAACAAAACTCGAGGCTCTTGGAAAATGG 253

     178 spArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGln 194
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     252 ACAGATTTCTGAACCGTTTCCACCTGGGCGAACCTGAAGCAAGCACCCAG 203

     195 PheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaProArgGl 211
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TTCATGACCCAGAACTACCAGGACTCACCAACCCTCCAGGCTCCCAGAGA 153

     211 uArgAlaSerGluProLysHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     152 AAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCGGAGATAGCTGCCTGC 104

seq_name: gb_htg13:AC055854

seq_documentation_block:
LOCUS       AC055854   179814 bp    DNA             HTG       25-MAY-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-459E5 map 8, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC055854
VERSION     AC055854.3  GI:8084680
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179814)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-459E5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179814)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7770648.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------ Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------ Project Information
                Center project name: L9765
                Center clone name: 459_E_5
            ------------ Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167541 bases at least Q40
                Consensus quality: 173729 bases at least Q30
                Consensus quality: 175997 bases at least Q20
                Insert size: 175000; agarose-fp
                Insert size: 177314; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1730: contig of 1730 bp in length
            *    1731 1830: gap of      100 bp
            *    1831     3230: contig of 1400 bp in length
            *    3231 3330: gap of      100 bp
            *    3331     5675: contig of 2345 bp in length
            *    5676 5775: gap of      100 bp
            *    5776     7874: contig of 2099 bp in length
            *    7875 7974: gap of      100 bp
            *    7975    10477: contig of 2503 bp in length
            *   10478 10577: gap of      100 bp
            *   10578    12553: contig of 1976 bp in length
            *   12554 12653: gap of      100 bp
            *   12654    16386: contig of 3733 bp in length
            *   16387 16486: gap of      100 bp
            *   16487    19715: contig of 3229 bp in length
            *   19716 19815: gap of      100 bp
            *   19816    23166: contig of 3351 bp in length
            *   23167 23266: gap of      100 bp
            *   23267    26855: contig of 3589 bp in length
            *   26856 26955: gap of      100 bp
            *   26956    31690: contig of 4735 bp in length
            *   31691 31790: gap of      100 bp
            *   31791    38834: contig of 7044 bp in length
            *   38835 38934: gap of      100 bp
            *   38935    46670: contig of 7736 bp in length
            *   46671 46770: gap of      100 bp
            *   46771    52635: contig of 5865 bp in length
            *   52636 52735: gap of      100 bp
            *   52736    59853: contig of 7118 bp in length
            *   59854 59953: gap of      100 bp
            *   59954    66887: contig of 6934 bp in length
            *   66888 66987: gap of      100 bp
            *   66988    75106: contig of 8119 bp in length
            *   75107 75206: gap of      100 bp
            *   75207    84076: contig of 8870 bp in length
            *   84077 84176: gap of      100 bp
            *   84177    91487: contig of 7311 bp in length
            *   91488 91587: gap of      100 bp
            *   91588    98734: contig of 7147 bp in length
            *   98735 98834: gap of      100 bp
            *   98835   107787: contig of 8953 bp in length
            *  107788 107887: gap of      100 bp
            *  107888   118191: contig of 10304 bp in length
            *  118192 118291: gap of      100 bp
```

```
          *   118292   134239: contig of 15948 bp in length
          *   134240 134339: gap of      100 bp
          *   134340   148401: contig of 14062 bp in length
          *   148402 148501: gap of      100 bp
          *   148502   163397: contig of 14896 bp in length
          *   163398 163497: gap of      100 bp
          *   163498   179814: contig of 16317 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179814
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-459E5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1730
                     /note="assembly_fragment"
     misc_feature    1831. .3230
                     /note="assembly_fragment"
     misc_feature    3331. .5675
                     /note="assembly_fragment"
     misc_feature    5776. .7874
                     /note="assembly_fragment"
     misc_feature    7975. .10477
                     /note="assembly_fragment"
     misc_feature    10578. .12553
                     /note="assembly_fragment"
     misc_feature    12654. .16386
                     /note="assembly_fragment"
     misc_feature    16487. .19715
                     /note="assembly_fragment"
     misc_feature    19816. .23166
                     /note="assembly_fragment"
     misc_feature    23267. .26855
                     /note="assembly_fragment"
     misc_feature    26956. .31690
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    31791. .38834
                     /note="assembly_fragment"
     misc_feature    38935. .46670
                     /note="assembly_fragment"
     misc_feature    46771. .52635
                     /note="assembly_fragment"
     misc_feature    52736. .59853
                     /note="assembly_fragment"
     misc_feature    59954. .66887
                     /note="assembly_fragment"
     misc_feature    66988. .75106
                     /note="assembly_fragment"
     misc_feature    75207. .84076
                     /note="assembly_fragment"
     misc_feature    84177. .91487
                     /note="assembly_fragment"
     misc_feature    91588. .98734
                     /note="assembly_fragment"
     misc_feature    98835. .107787
                     /note="assembly_fragment"
     misc_feature    107888. .118191
                     /note="assembly_fragment"
     misc_feature    118292. .134239
                     /note="assembly_fragment"
     misc_feature    134340. .148401
                     /note="assembly_fragment"
     misc_feature    148502. .163397
                     /note="assembly_fragment"
     misc_feature    163498. .179814
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
BASE COUNT    43709 a  43721 c  45547 g  44332 t   2505 others
ORIGIN
```

```
alignment_scores:
            Quality:  262.50            Length:    137
              Ratio:  3.052               Gaps:      6
 Percent Similarity:  62.774   Percent Identity: 51.095

alignment_block:
 US-09-215-435-225 x AC055854/rev   ..

 Align seg 1/1  to reverse of: AC055854  from: 1  to: 179814

      96 AspProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHi 112
         :::|||:::|||||||      :::   |||
   76358 GAGCCGGAAGCTCCCCCAGCCACTACGCCC..................... 76329

     112 sTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleG 129
                                                          :
   76328 .............................................C 76328

     129 lnGlyGlnGluLeuSer.AlaTyrGlnAlaProSer......ProPro.. 142
         ::|||:::::::::::: |||   :::||||||:::      ||||||
   76327 ATGGCCATCGCATCACCTGCTGCCTCGGCTCCCACGAGCCGGCCTCCCTC 76278.

     143 ....AlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnG 158
             |||   ::::::|||:::   :::      |||
   76277 TCAGGCCTCTGCCAGCTTTCAAGTCCATGGCTCTTTCTCACTGCCTCCCT 76228

     158 luGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySer 174
           |||      :::   :::||||||                  ||||||
   76227 GTGGT......CTGTTTGTCCTTCCA.................GGCTCT 76202

     175 TrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAl 191
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   76201 TGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACCTGAAGC 76152

     191 aSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnA 208
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   76151 AAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCCTCCAGG 76102

     208 laProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaGluIle 224
         ||||||||   ||||||||||||||||||||   ||||||||||||||||
   76101 CTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.AACCAGGCGGAGATA 76053

     225 AlaAlaCys 227
         |||||||||
   76052 GCTGCCTGC 76044

seq_name: gb_htg12:AC037441

seq_documentation_block:
LOCUS       AC037441   162795 bp    DNA             HTG       03-MAR-2001
DEFINITION  Homo sapiens clone RP11-87E22, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC037441
VERSION     AC037441.3  GI:13184260
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-87E22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
```

```
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:8705171.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9203
                Center clone name: 87_E_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 81% of reads
                Sequencing vector: Plasmid; n/a; 19% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 157025 bases at least Q40
                Consensus quality: 160134 bases at least Q30
                Consensus quality: 161093 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 161795; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15243: contig of 15243 bp in length
            *    15244 15343: gap of      100 bp
            *    15344    18897: contig of 3554 bp in length
            *    18898 18997: gap of      100 bp
            *    18998    25211: contig of 6214 bp in length
            *    25212 25311: gap of      100 bp
            *    25312    34838: contig of 9527 bp in length
            *    34839 34938: gap of      100 bp
            *    34939    45351: contig of 10413 bp in length
            *    45352 45451: gap of      100 bp
            *    45452    57259: contig of 11808 bp in length
            *    57260 57359: gap of      100 bp
            *    57360    67990: contig of 10631 bp in length
            *    67991 68090: gap of      100 bp
            *    68091    83399: contig of 15309 bp in length
            *    83400 83499: gap of      100 bp
            *    83500   105395: contig of 21896 bp in length
            *   105396 105495: gap of      100 bp
            *   105496   129959: contig of 24464 bp in length
            *   129960 130059: gap of      100 bp
            *   130060   162795: contig of 32736 bp in length.
  FEATURES             Location/Qualifiers
       source          1. .162795
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="RP11-87E22"
                       /clone_lib="RPCI-11 Human Male BAC"
       misc_feature    1. .15243
                       /note="assembly_fragment
                       clone_end:SP6
                       vector_side:left"
       misc_feature    15344. .18897
                       /note="assembly_fragment"
       misc_feature    18998. .25211
                       /note="assembly_fragment"
       misc_feature    25312. .34838
                       /note="assembly_fragment"
       misc_feature    34939. .45351
                       /note="assembly_fragment"
       misc_feature    45452. .57259
                       /note="assembly_fragment"
       misc_feature    57360. .67990
                       /note="assembly_fragment"
       misc_feature    68091. .83399
                       /note="assembly_fragment"
       misc_feature    83500. .105395
                       /note="assembly_fragment"
       misc_feature    105496. .129959
                       /note="assembly_fragment"
       misc_feature    130060. .162795
                       /note="assembly_fragment
                       clone_end:T7
                       vector_side:right"
  BASE COUNT    42087 a  39347 c  38304 g  42057 t   1000 others
  ORIGIN

  alignment_scores:
              Quality:  246.00              Length:    239
                Ratio:  2.103                 Gaps:     15
   Percent Similarity:  48.954   Percent Identity: 37.238

  alignment_block:
   US-09-215-435-225 x AC037441/rev   ..

   Align seg 1/1  to reverse of: AC037441  from: 1  to: 162795

         1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
           ||||||||||||||||||||||||||||||||||||||||||||||||||
    159302 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 159253

        17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
           ||||||||||||||||||||||||||||||||||||||||||||||||||
    159252 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 159203

        34 laLeuLeuAspGluAspThrLeuPheCysGln.....GlyLeu 46
           ||||||||||||||||||||||||||||||||     ||||||
    159202 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGTAAGTCAGGGCAGGCTTG 159153

        47 GluValPheTyrProGluLeuGlyAsnIleGlyCysLysValValProAs 63
           |||:::        |||
    159152 GAGGCGGGAGGCACAGAGGCT..................................... 159132

        63 pCysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValL 80
                         ||||||
    159131 .........ACCAGGCAG..................................... 159123

        80 ysPheProGlyAlaValAspGlyAlaThr...TyrIleLeuValMetVal 95
                   |||||||||         ||||||
    159122 ..............GTAGATGGCAGGGAGGGGTATATC.........CAG 159096

        96 AspProAspAlaProSerArgAlaGluPro.....ArgGln.. 107
           :::   :::|||||||||      ||||||      ||||||
    159095 AACCAAGAAGCACCATCT...CCTGAGCCTGATTTTATGGGAAGGCAGAA 159049

       108 .......ArgPheTrpArgHis......TrpLeuValThrAspIleLysG 120
                 ::::::|||||         |||   ::::::
```

```
  159048 CACGTGTCAGATCTGGAGACCACCTTGATGGAGGATGGCA..... 159009

     120 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 136
                ||||::|||||||   ::::::::::::   |||   |||::: 
  159008 ........CTGAGAAAGGGA...TTTAGAAACAGACTCCTTCCTGCCTTC 158970

     137 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 153
         |||            |||:::   :::|||:::      :::
  158969 CAGAATGCACCTCTCCCAGGACCAGGAGGGCTGTGTATGTGG........ 158928

     153 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 170
                                  |||   |||         :::|
  158927 ............................TCGGCTCTGCTTGGGCCCTCCA 158906

     170 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 186
         ||      ||||||                                |||
  158905 AACCTGGGGGGTCT................................TTG 158889

     187 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 203
                 ||||||:::|||      :::   ::::::   ::::::
  158888 TGGGGGCAGGAGGCTGAGACC......CTCAGAGAGAGACCCAGAGAGAG 158845

     203 rProThrLeuGlnAla............ProArgGluArgAlaSerGluP 216
          |||:::|||   |||             |||      |||:::
  158844 ACCATCCCTCCCTGCAGCCTGAATGCCCCCTCTCCCCAGAACCCCAACAA 158795

     216 roLysHisLysAsnGln 221
           :::|||   |||:::
  158794 AGAGGCATTTGAATGAA 158778

seq_name: gb_htg17:AC084081

seq_documentation_block:
LOCUS       AC084081   169645 bp    DNA             HTG       11-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-89M8 map 8, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC084081
VERSION     AC084081.3  GI:13273380
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-89M8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2001 this sequence version replaced gi:12583824.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11307
                Center clone name: 89_M_8
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162624 bases at least Q40
                Consensus quality: 165742 bases at least Q30
                Consensus quality: 167046 bases at least Q20
                Insert size: 169000; agarose-fp
                Insert size: 168045; sum-of-contigs
                Quality coverage: 5.9 in Q20 bases; agarose-fp
                Quality coverage: 5.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      335: contig of 335 bp in length
            *     336 435: gap of      100 bp
            *     436     2289: contig of 1854 bp in length
            *    2290 2389: gap of      100 bp
            *    2390     4997: contig of 2608 bp in length
            *    4998 5097: gap of      100 bp
            *    5098    64677: contig of 59580 bp in length
            *   64678 64777: gap of      100 bp
            *   64778    68151: contig of 3374 bp in length
            *   68152 68251: gap of      100 bp
            *   68252    72655: contig of 4404 bp in length
            *   72656 72755: gap of      100 bp
            *   72756    78893: contig of 6138 bp in length
            *   78894 78993: gap of      100 bp
            *   78994    84664: contig of 5671 bp in length
            *   84665 84764: gap of      100 bp
            *   84765    90853: contig of 6089 bp in length
            *   90854 90953: gap of      100 bp
            *   90954    98425: contig of 7472 bp in length
            *   98426 98525: gap of      100 bp
            *   98526   106371: contig of 7846 bp in length
            *  106372 106471: gap of      100 bp
            *  106472   115673: contig of 9202 bp in length
            *  115674 115773: gap of      100 bp
            *  115774   127138: contig of 11365 bp in length
            *  127139 127238: gap of      100 bp
            *  127239   138436: contig of 11198 bp in length
            *  138437 138536: gap of      100 bp
            *  138537   149766: contig of 11230 bp in length
            *  149767 149866: gap of      100 bp
            *  149867   167340: contig of 17474 bp in length
            *  167341 167440: gap of      100 bp
            *  167441   169645: contig of 2205 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-89M8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .335
                     /note="assembly_fragment
```

```
                clone_end:SP6
                vector_side:left"
    misc_feature    436. .2289
                /note="assembly_fragment"
    misc_feature    2390. .4997
                /note="assembly_fragment"
    misc_feature    5098. .64677
                /note="assembly_fragment"
    misc_feature    64778. .68151
                /note="assembly_fragment"
    misc_feature    68252. .72655
                /note="assembly_fragment"
    misc_feature    72756. .78893
                /note="assembly_fragment"
    misc_feature    78994. .84664
                /note="assembly_fragment"
    misc_feature    84765. .90853
                /note="assembly_fragment"
    misc_feature    90954. .98425
                /note="assembly_fragment"
    misc_feature    98526. .106371
                /note="assembly_fragment"
    misc_feature    106472. .115673
                /note="assembly_fragment"
    misc_feature    115774. .127138
                /note="assembly_fragment"
    misc_feature    127239. .138436
                /note="assembly_fragment"
    misc_feature    138537. .149766
                /note="assembly_fragment"
    misc_feature    149867. .167340
                /note="assembly_fragment"
    misc_feature    167441. .169645
                /note="assembly_fragment
                clone_end:T7
                vector_side:right"
BASE COUNT    44253 a  41074 c  39720 g  42960 t   1638 others
ORIGIN


alignment_scores:
            Quality:  246.00              Length:    231
              Ratio:  2.067                Gaps:     12
 Percent Similarity:  51.515   Percent Identity: 36.364

alignment_block:
 US-09-215-435-225 x AC084081/rev   ..

 Align seg 1/1  to reverse of: AC084081  from: 1  to: 169645

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  49659 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 49610

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  49609 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 49560

     34 laLeuLeuAspGluAspThrLeuPheCysGln.....GlyLeu 46
        ||||||||||||||||||||||||||||||||     ||||||
  49559 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGTAAGTCAGGGCAGGCTTG 49510

     47 GluValPheTyrProGluLeuGlyAsnIleGlyCysLysValValProAs 63
        |||:::          |||:::   :::   :::   |||
  49509 GAGGCGGGA.....GGCACAGAGGCTACCAGGCAGGTAGATGG 49472

     63 pCysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValL 80
          :::::|||   |||:::
  49471 CAGGGAGGGGTATATCCAGAAC..................... 49450

     80 ysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAsp 96
                  :::|||||||                      :::
  49449 ..............CAAGGAGCACCA.....TCTCCTGAG 49429

     97 ProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis.. 112
        ||||||      :::|||   :::      |||   :::|||||
  49428 CCTGATTTTATGGGAAGGCAGAACACGTGTCAG...ATCTGGAGACCACC 49382

    113 ....TrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysI 128
            |||   ::::::                  |||:::||||||   :
  49381 TTGATGGAGGATGGCA.............CTGAGAAAGGGA...T 49353

    128 leGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHis 144
        ::::::::::   |||   |||:::|||           |||:::
  49352 TTAGAAACAGACTCCTTCCTGCCTTCCAGAATGCACCTCTCCCAGGACCA 49303

    145 SerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysVa 161
        :::|||:::     :::
  49302 GGAGGGCTGTGTATGTGG.............................. 49285

    161 lIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetA 178
           |||   |||         :::|||      ||||||
  49284 ....TCGGCTCTGCTTGGGCCCTCCAAACCTGGGGGGTCT......... 49249

    178 spArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGln 194
                                  |||         ||||||:::|||
  49248 .....................TTGTGGGGGCAGGAGGCTGAGACC... 49225

    195 PheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAla........ 208
           :::   ::::::   ::::::   |||:::|||   |||
  49224 ...CTCAGAGAGAGACCCAGAGAGAGACCATCCCTCCCTGCAGCCTGAAT 49178

    209 ....ProArgGluArgAlaSerGluProLysHisLysAsnGln 221
           |||       |||:::         :::|||   |||:::
  49177 GCCCCCTCTCCCCAGAACCCCAACAAAGAGGCATTTGAATGAA 49135

seq_name: gb_ro2:RNPBP

seq_documentation_block:
LOCUS       RNPBP        1037 bp    mRNA            ROD       17-FEB-1997
DEFINITION  R.norvegicus mRNA for phosphatidylethanolamine-binding protein.
ACCESSION   X71873
VERSION     X71873.1  GI:510338
KEYWORDS    epididymal protein; pbp gene; phosphatidylethanolamine-binding
            protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   Perry,A.C., Hall,L., Bell,A.E. and Jones,R.
  TITLE     Sequence analysis of a mammalian phospholipid-binding protein from
            testis and epididymis and its distribution between spermatozoa and
            extracellular secretions
  JOURNAL   Biochem. J. 301 (Pt 1), 235-242 (1994)
  MEDLINE   94311839
REFERENCE   2  (bases 1 to 1037)
  AUTHORS   Hall,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1993) L. Hall, University of Bristol, Dept of
            Biochemistry, School of Medical Sciences, University Walk, Bristol
            BS8 1TD, UK
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /dev_stage="adult male"
                     /tissue_type="epididymis and liver"
                     /clone_lib="cDNA"
                     /clone="18 clones"
     5'UTR           1. .28
     gene            29. .592
                     /gene="PBP"
     CDS             29. .592
                     /gene="PBP"
```

```
                    /codon_start=1
                    /product="phosphatidylethanolamine-binding protein"
                    /protein_id="CAA50708.1"
                    /db_xref="GI:510339"
                    /db_xref="SWISS-PROT:P31044"
                    /translation="MAADISQWAGPLSLQEVDEPPQHALRVDYGGVTVDELGKVLTPT
                    QVMNRPSSISWDGLDPGKLYTLVLTDPDAPSRKDPKFREWHHFLVVNMKGNDISSGTV
                    LSEYVGSGPPKDTGLHRYVWLVYEQEQPLNCDEPILSNKSGDNRGKFKVESFRKKYHL
                    GAPVAGTCFQAEWDDSVPKLHDQLAGK"
     3'UTR          590. .1037
     variation      766
                    /note="in 6 clones"
                    /replace="t"
BASE COUNT      244 a    299 c    290 g    204 t
ORIGIN


alignment_scores:
           Quality:  237.50            Length:    185
             Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225 x RNPBP   ..

 Align seg 1/1  to: RNPBP  from: 1  to: 1037

      52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
         ||||||||||:::::        ::::::|||               ||
     134 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 174

      68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
         |   :::      ||||||                          |||
     175 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 203

      85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
            |||      |||   ||||||::::::||||||||||||||||||
     204 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 253

     102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     254 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 303

     118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
         :||||||   |||:::::::::|||   :::             ||||||
     304 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 341

     135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 151
           |||   :::   :::||||||      :::|||:::||||||||||
     342 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 391

     152 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 164
         ::::::||||||   ||||||      :::                 ||
     392 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 441

     164 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 181
         |      |||::::::   :::||||||:::::|||:::::   |||
     442 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 491

     181 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 197
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     492 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 541

     198 GlnAsnTyrGlnAspSer.............................. 203
         :::   ::::::||||||
     542 GAG...TGGGATGACTCTGTGCCCAAGCTGCACGATCAGCTGGCTGGGAA 588

     204 ................................ProThrLeuGlnAlaP 209
                                        ||||||:::|||:::
     589 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 638

     209 roArg 210
           |||
```

```
     639 GTCGT 643

seq_name: gb_pat2:E05646

seq_documentation_block:
LOCUS       E05646       1047 bp    RNA             PAT       29-SEP-1997
DEFINITION  cDNA encoding rat hippocampal cholinergic neurostimulating
            peptide,HCNP.
ACCESSION   E05646
VERSION     E05646.1  GI:2173833
KEYWORDS    JP 1993268966-A/1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   Tofuji,N., Tojo,S., Kojima,S., Ueki,Y., Nishihara,N., Fukushima,N.,
            Irie,T., Ono,K., Agui,H. and Kojika,Y. .
  TITLE     NEUROTROPHIC FACTOR-RELATED POLYPEPTIDE AND GENE USEFUL THEREFOR
  JOURNAL   Patent: JP 1993268966-A 1 19-OCT-1993;
            SUMITOMO PHARMACEUT CO LTD, KOJIKA YUKIO, YAMAMOTO MASAHIKO
COMMENT     OS   Rattus norvegicus (rat)
            PN   JP 1993268966-A/1
            PD   19-OCT-1993
            PF   25-APR-1992 JP 1992131703
            PR   27-APR-1991 JP 91P   124688
            PI   TOFUJI NAOKI, TOJO SHINICHIRO, KOJIMA SHINICHI, UEKI YASUYUKI,
            PI   NISHIHARA NORIO, FUKUSHIMA NOBUYUKI, IRIE TSUNEMASA, PI   ONO
            KEIICHI,
            PI   AGUI HIDEO, KOJIKA YUKIO
            PC   C12N15/12,A61K37/02,A61K37/02,C07K7/06,C07K13/00,C12N1/21, PC
             C12N5/10,
            PC   (C12N1/21,C12R1:19),(C12N5/10,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: tissue_type=hippocampus of brain;
            CC   *source: clone=AO10-12;
            CC   Feature is identified by similarity;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           <1. .25
            FT   CDS             26. .589
            FT                   /product='hippocampal cholinergic  FT
                 neurostimulating peptide'
            FT   3'UTR           590. .1047
            FT   polyA_signal    987. .992
            FT   polyA_site      1008. .1047
            FT   mutation        replace(427. .429,'AAG')
            FT                   /note='hippocampal cholinergic  FT
                  neurostimulating peptide
            FT                   derived from clone AO1-1':
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
BASE COUNT      257 a    297 c    289 g    204 t
ORIGIN


alignment_scores:
           Quality:  237.50            Length:    185
             Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225 x E05646   ..

 Align seg 1/1  to: E05646  from: 1  to: 1047

      52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
```

```
     |||||||||:::::::         :::::::|||                  ||
   131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

    68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
       |   :::      ||||||                          |||
   172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

    85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
         |||      |||   ||||||:::::|||||||||||||||||||||
   201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

   102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
       |||   ::|||:::   |||   |||:::|||:::||||||:::::::
   251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

   118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
       :||||||   |||:::::::::|||   :::             ||||||
   301 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 338

   135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 151
         |||   :::   :::|||||||      :::|||:::|||||||||
   339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

   152 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 164
       ::::::|||||||   ||||||      :::                ||
   389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

   164 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 181
       |      |||::::::   :::||||||:::::::|||::::::   |||
   439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

   181 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 197
         :::::::::|||||||||   |||   |||:::|||   |||::::::
   489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

   198 GlnAsnTyrGlnAspSer.............................. 203
       :::   ::::::|||||||
   539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585

   204 ................................ProThrLeuGlnAlaP 209
                                       ||||||:::|||:::
   586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

   209 roArg 210
         |||
   636 GTCGT 640

seq_name: gb_ro2:RNPEABP

seq_documentation_block:
LOCUS       RNPEABP      1075 bp    mRNA            ROD       25-AUG-1995
DEFINITION  R.norvegicus phosphatidylethanolamine binding protein mRNA.
ACCESSION   X75253
VERSION     X75253.1  GI:406291
KEYWORDS    phosphatidylethanolamine-binding protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1075)
  AUTHORS   Tohdoh,N., Tojo,S., Agui,H. and Ojika,K.
  TITLE     Sequence homology of rat and human HCNP precursor proteins, bovine
            phosphatidylethanolamine-binding protein and rat 23-kDa protein
            associated with the opioid-binding protein
  JOURNAL   Brain Res. Mol. Brain Res. 30 (2), 381-384 (1995)
  MEDLINE   95364631
REFERENCE   2  (bases 1 to 1075)
  AUTHORS   Tohdoh,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1993) N. Tohdoh, Sumitomo Pharmaceuticals Co.,
            Ltd., 1-98, Kasugade-Naka 3-chome, Konohanaku, Osaka, 554, JAPAN
FEATURES             Location/Qualifiers
     source          1. .1075
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /germline
                     /tissue_type="hippocampus"
                     /clone_lib="lambda gt11"
                     /clone="pA01-1, pA010-1, pA62"
     mRNA            1. .1075
     CDS             54. .617
                     /codon_start=1
                     /product="phosphatidylethanolamine binding protein"
                     /protein_id="CAA53032.1"
                     /db_xref="GI:406292"
                     /db_xref="SWISS-PROT:P31044"
                     /translation="MAADISQWAGPLSLQEVDEPPQHALRVDYGGVTVDELGKVLTPT
                     QVMNRPSSISWDGLDPGKLYTLVLTDPDAPSRKDPKFREWHHFLVVNMKGNDISSGTV
                     LSEYVGSGPPKDTGLHRYVWLVYEQEQPLNCDEPILSNKSGDNRGKFKVESFRKKYHL
                     GAPVAGTCFQAEWDDSVPKLHDQLAGK"
     polyA_signal    1015. .1020
     polyA_signal    1036. .1075
BASE COUNT      261 a    305 c    298 g    211 t
ORIGIN


alignment_scores:
            Quality:  237.50              Length:    185
              Ratio:  2.284                 Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225 x RNPEABP   ..

 Align seg 1/1  to: RNPEABP  from: 1  to: 1075

     52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
        |||||||||:::::         ::::::|||                ||
    159 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 199

     68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
        |   :::      ||||||                          |||
    200 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 228

     85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
          |||      |||   ||||||:::::|||||||||||||||||||||
    229 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 278

    102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
        |||   :::|||:::   |||   |||:::|||:::||||||:::::::
    279 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 328

    118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
        :||||||   |||:::::::::|||   :::             ||||||
    329 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 366

    135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 151
          |||   :::   :::|||||||      :::|||:::|||||||||
    367 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 416

    152 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 164
        ::::::|||||||   ||||||      :::                ||
    417 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 466

    164 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 181
        |      |||::::::   :::||||||:::::::|||::::::   |||
    467 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 516

    181 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 197
          :::::::::|||||||||   |||   |||:::|||   |||::::::
    517 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 566

    198 GlnAsnTyrGlnAspSer.............................. 203
        :::   ::::::|||||||
```

```
     567 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 613

     204 ..............................ProThrLeuGlnAlaP 209
                                       ||||||:::|||::: 
     614 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 663

     209 roArg 210
         |||
     664 GTCGT 668

seq_name: gb_in4:OVD1PROTN

seq_documentation_block:
LOCUS       OVD1PROTN     611 bp    mRNA            INV       05-NOV-1996
DEFINITION  O.volvulus mRNA for d1 protein.
ACCESSION   X87991
VERSION     X87991.1  GI:1143526
KEYWORDS    D1 gene; D1 protein.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Erttmann,K.D. and Gallin,M.Y.
  TITLE     Onchocerca volvulus: identification of cDNAs encoding a putative
            phosphatidyl-ethanolamine-binding protein and a putative partially
            processed mRNA precursor
  JOURNAL   Gene 174 (2), 203-207 (1996)
  MEDLINE   97045813
REFERENCE   2  (bases 1 to 611)
  AUTHORS   Erttmann,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1995) K.D. Erttmann, Bernhard Nocht Inst for
            Tropical Med., Bernhard-Nocht-Str. 74, 20359 Hamburg, FRG
COMMENT     Related sequences M27807, U05980, U05981, U05982, U05984, U05985,
            U05243, U05244.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Onchocerca volvulus"
                     /db_xref="taxon:6282"
                     /dev_stage="adult"
                     /clone_lib="lambda ZAP"
                     /cell_line="E.Coli"
     gene            1. .459
                     /gene="d1"
     CDS             <1. .459
                     /gene="d1"
                     /codon_start=1
                     /product="D1 protein"
                     /protein_id="CAA61244.1"
                     /db_xref="GI:1143527"
                     /db_xref="SWISS-PROT:P54186"
                     /translation="VVSTSPTKLVNVSYNNLTVNLGNELTPTQVKNQPTKVSWDAEPG
                     ALYTLVMTDPDAPSRKNPVFREWHHWLIINISGQNVSSGTVLSDYIGSGPPKGTGLHR
                     YVFLVYKQPGSITDTQHGGNRPNFKVMDFANKHHLGNPVAGNFFQAKHED"
BASE COUNT      196 a    114 c    123 g    178 t
ORIGIN


alignment_scores:
            Quality:  234.50            Length:    152
              Ratio:  2.393               Gaps:      6
 Percent Similarity:  64.474   Percent Identity: 36.842

alignment_block:
 US-09-215-435-225 x OVD1PROTN   ..

 Align seg 1/1  to: OVD1PROTN  from: 1  to: 611

      55 AsnIleGlyCysLysValValPro...AspCysAsnAsnTyrArgGlnLy 70
         |||:::|||   :::::::::|||   :::   :::|||        ||
      58 AATCTGGGCAATGAACTTACGCCGACGCAGGTAAAGAATCAGCCGACAAA 107

      70 sIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAspG 87
         |:::   ||||||                              |||      |
     108 AGTA...TCATGG.....................GATGCGGAACCTG 130

      87 lyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAla 103
         |||||   |||   |||||||||:::|||||||||||||||||||||
     131 GAGCCTTATATACGCTCGTTATGACTGATCCGGACGCACCATCTCGAAAA 180

     104 GluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGl 120
         :::|||      |||   |||:::||||||||:::   :::|||:::||
     181 AACCCCGTATTCAGAGAGTGGCACCATTGGTTGATAATTAATATTTCTGG 230

     120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrG 137
         |   :::::::::::|||  :::            ||||||   |||
     231 ACAAAATGTTAGCAGTGGCACAGTG...........TTATCTGATTATA 268

     137 lnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePhe 153
           :::   :::||||||      :::|||:::|||||||||   |||:::
     269 TTGGATCAGGTCCACCAAAAGGCACAGGACTTCATCGTTATGTATTCTTG 318

     154 ValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLy 170
         ||||||   |||   |||::::::   :::         :::::::::
     319 GTTTATAAACAACCTGGAAGTATCACGGAT......ACTCAACATGGCGG 362

     170 sThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuG 187
          :::|||   ::::::|||:::      |||   |||:::   |||||||
     363 AAATCGCCCAAATTTCAAAGTTATGGATTTTGCAAACAAACATCACTTGG 412

     187 lyGluProGluAlaSerThrGlnPheMetThrGlnAsn.........Tyr 200
         ||:::|||   |||::::::   |||:::::::::::::         |||
     413 GAAATCCAGTTGCCGGGAACTTCTTCCAGGCTAAACATGAAGACTAATAT 462

     201 GlnAsp 202
         :::|||
     463 GAAGAC 468

seq_name: gb_in3:CELF40A3

seq_documentation_block:
LOCUS       CELF40A3    41155 bp    DNA             INV       02-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid F40A3.
ACCESSION   AF016423
VERSION     AF016423.1  GI:2291195
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41155)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 41155)
  AUTHORS   Geisel,C., Bradshaw,H. and Keppler,D.
  TITLE     The sequence of C. elegans cosmid F40A3
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 41155)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1997)
COMMENT     Submitted by:
```

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C37C3, 200 bp overlap;3' lies in a gap. Actual start of this cosmid is at base position 197 of CELF40A3; actual end is at 41155 of CELF40A3

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .41155
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="F40A3"
     gene            complement(156. .1605)
                     /gene="F40A3.5"
     CDS             complement(join(156. .216,268. .662,714. .787,864. .981,
                     1094. .1281,1357. .1431,1476. .1605))
                     /gene="F40A3.5"
                     /note="contains similarity to a protein kinase domain"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65324.1"
                     /db_xref="GI:2291201"
                     /translation="MELDHTPPPSVLNDNCSASYMTPYATVIAMSGLYLLAIFYFCKK
                     SKKMCQPMSDSIYPYQKRLTQLERELFKDVRAHSVFFAELKNYLIDEESIEVDDFQIG
                     QTADGFIFRGGVFPKTRNRFNAKVTTAVKISFPIVSKSISLLEDALRLSKLDHPNLIR
                     LLAVSQLSFTVFRPMIALEWLPGGTLADYFQFKVREKDDSERSPIQLKDMLSILYQVS
                     QALKYIHSQLDEFGQELTHGRIFTRNVLVTEPDLRKCEVKLGDFGDAPMGLEYSTPII
                     AYMPPEILCCAERIPPHRPENDVWMFGVFIWECLTLGAQPHFRKSVEEIKKSFRLPDR
                     GLSCPPTCPLDV"
     gene            6740. .7594
                     /gene="F40A3.4"
     CDS             join(6740. .6827,7331. .7394,7450. .7594)
                     /gene="F40A3.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65323.1"
                     /db_xref="GI:2291200"
                     /translation="MSRMDSYSSSIREEQKAILNGLLGILFADGIIYDNYDTTRKIHQ
                     TVDGIYRMVVALLLIVIALIIIAVFLICFFHCRRANNEERSPLHSNRKSHPIEI"
     gene            8506. .9283
                     /gene="F40A3.3"
     CDS             join(8506. .8600,8667. .8781,8828. .9283)
                     /gene="F40A3.3"
                     /note="Similar to phosphatidylethanolamine binding
                     protein; coded for by C. elegans cDNA yk106a12.5; coded
                     for by C. elegans cDNA yk106a12.3"
                     /codon_start=1
                     /protein_id="AAB65322.1"
                     /db_xref="GI:2291199"
                     /translation="MVVLVTRSLLPALFFASRAPFAAATTSARFQRGLATMAAEAFTK
                     HEVIPDVLASNPPSKVVSVKFNSGVEANLGNVLTPTQVKDTPEVKWDAEPGALYTLIK
                     TDPDAPSRKEPTYREWHHWLVVNIPGNDIAKGDTLSEYIGAGPPPKTGLHRYVYLIYK
                     QSGRIEDAEHGRLTNTSGDKRGGWKAADFVAKHKLGAPVFGNLFQAEYDDYVPILNKQ
                     LGA"
     gene            9824. .12167
                     /gene="F40A3.2"
     CDS             join(9824. .9917,10735. .11058,11837. .12034,12091. .12167)
                     /gene="F40A3.2"
                     /note="contains similarity to EGF-like domains"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65320.1"
                     /db_xref="GI:2291197"
                     /translation="MLRRILLSAILIVLVACVLRKKKPCINGTPEGDRCYCIEGWTGT
                     LCHRKMYCSGFERLQNGSCIECELGWAGTDCDIIDCHGNGMPNYDLTECTCTVPYSGK
                     YCEIADTKDIYKWYNNFTSSIGPIGILTIIPLALIYMCCEHFAKKRQLRRVEQHLNGV
                     YITTGRKQVDKMLVKGLLEASDDEKSVKSVKIAADSGKEDKDKEEEEVIKMRNSATTS
                     EKKALLADPADM"
     gene            complement(23146. .24752)
                     /gene="F40A3.6"
     CDS             complement(join(23146. .23311,23512. .23597,23644. .23841,
                     23885. .24052,24594. .24752))
                     /gene="F40A3.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65325.1"
                     /db_xref="GI:2291202"
                     /translation="MSEYLGWHNRKPAAGSRGNSSFDRDIVPGGAHTKSYESYVKNIF
                     HPTNRLRAKDKDQLVRDHGLSATSMLRAEREFGRSALDDKYDREHFFEKMPLESLYAE
                     LGLKNPAASNYRSFQVYEPKSSSEYFAASYEAERERRKKDLERELASYDKEDYEDFTG
                     EPKKNLYSSISSYRPPLSSYKSREDEEYEKNNRKTLIDYKDFDSSPLSPYLYQSHPYV
                     RSQDSRIIGSNFVQLTSRPKDKFISKIEETLAMVRDMPRY"
     gene            24968. .26143
                     /gene="F40A3.1"
     CDS             join(24968. .25107,25199. .25454,25498. .25634,26041. .26143)
                     /gene="F40A3.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65321.1"
                     /db_xref="GI:2291198"
                     /translation="MSEIVRLPEETQETEMPLHHANRAFGARKQDAGWMAKNEGKTFR
                     VVIKNIEIFQINDWELIGSRTYSWKGNEVLRCKHWDKVAIRWHLSTCTEHGVCDGGTN
                     VDYATRAQNWTEMPVWRNAFTYVQIPGQLQKTKFLRLGLTGYDSRKTQLEAFSPLKID
                     VDWRWQVTRDGCNNFVVRMSAMSMINEKMRPKLERFCRLFDKDHFVDCTNL"
     gene            complement(26520. .30751)
                     /gene="F40A3.7"
     CDS             complement(join(26520. .26710,27914. .28026,28197. .28358,
                     28415. .28609,29618. .29761,29814. .29935,30397. .30616,
                     30669. .30751))
                     /gene="F40A3.7"
                     /note="conatains similarity to seven-transmembrane
                     receptors"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65319.1"
                     /db_xref="GI:2291196"
                     /translation="MSANFSHLCLTDEQQRMSGTNLEYNLQRYFYPALAIFGILGNVL
                     NLTVLLNRSMRSRANSFLAVLAFADIIFLFLLFPNILANYSFFTFNWYFRWFYLHSKV
                     HLISLANWCSSVAHWCVIAVCGDRLLGIQNPLYARATWRWWKLPLVTTIIVFTCGLLT
                     CYQHFEYFCLVRSYCRETQLYSRCLPVNAEKWFGHRPNPFSQRYQSFIAICKLAHIFL
                     MIILPIILLLFLNLTLLWALRKRQKHLSIGKDFNADRRQNDVHMQKTEHRVTLTVTFI
                     VTMFTLTNGPSALVHLVMYATHEELYDLTMISSTLVICGKASNFILFCLGSKHFRLRL
                     LKLTRKKINRKIDSITGSLVHTTKLSTASSRRTSMPIQEKKRSSSCAPLVAKNRALST
                     VETIYQPLLNGNRKGDF"
BASE COUNT    13599 a    7323 c    6798 g   13435 t
ORIGIN


alignment_scores:
          Quality:  230.00              Length:    159
```

```
            Ratio:  2.447              Gaps:      7
 Percent Similarity:  59.119   Percent Identity: 37.107

alignment_block:
 US-09-215-435-225 x CELF40A3   ..

 Align seg 1/1  to: CELF40A3  from: 1  to: 41155

      49 PheTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAs 65
         |||      :::||||||||||:::        ::::::|||
    8822 TTTCAGGCTAATCTTGGAAACGTT.........CTCACTCCAACC..... 8857

      65 nAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPheP 82
            :::   |||   |||           |||   ||||||:::
    8858 .......CAAGTAAAGGATACT...........CCAGAAGTAAAATGG. 8887

      82 roGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAsp 98
           |||      ||||||   |||   |||:::   :::||||||||||
    8888 ..GATGCTGAGCCAGGAGCACTCTACACTCTCATCAAGACGGATCCAGAT 8935

      99 AlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuVa 115
         ||||||||||||   ||||||      |||   |||:::||||||||||||
    8936 GCTCCATCTCGCAAAGAGCCAACCTACCGAGAGTGGCATCACTGGCTCGT 8985

     115 lThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnG 132
         |:::::|||   |||   |||:::   |||             |||:::
    8986 CGTCAACATTCCAGGAAACGACATTGCCAAA...........GGAGACA 9023

     132 luLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHis 148
           ||||||   |||   :::   :::||||||      :::|||:::|||
    9024 CTCTTTCCGAGTACATCGGCGCCGGACCACCACCAAAAACTGGACTCCAT 9073

     149 ArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer..... 163
         ||||||   ::::::::::|||   |||:::|||::::::   :::
    9074 CGTTACGTTTACCTGATCTACAAGCAATCTGGACGCATCGAAGATGCCGA 9123

     164 .........LeuLeuProLysGluAsnLysThrArgGlySerTrpLysM 177
                  |||         ::::::      ||||||:::||||||
    9124 GCACGGACGTCTAACCAACACTTCCGGAGACAAAAGAGGAGGATGGAAGG 9173

     177 etAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThr 193
               |||:::   :::      ||||||   |||      ::::::
    9174 CTGCCGATTTCGTCGCAAAGCACAAGCTCGGTGCTCCAGTCTTCGGAAAT 9223

     194 GlnPheMetThrGlnAsnTyrGlnAsp 202
         |||::::::::::   |||:::|||
    9224 CTGTTCCAGGCCGAG...TACGACGAC 9247

seq_name: gb_htg1:AC006915

seq_documentation_block:
LOCUS       AC006915   134095 bp    DNA             HTG       24-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y97E10y, *** SEQUENCING IN PROGRESS
            ***, 1 unordered pieces.
ACCESSION   AC006915
VERSION     AC006915.1  GI:4263433
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 134095)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134095)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   134095: contig of 134095 bp in length.
FEATURES             Location/Qualifiers
     source          1. .134095
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y97E10y"
BASE COUNT    43230 a  24499 c  23702 g  42664 t
ORIGIN


alignment_scores:
            Quality:  230.00             Length:    159
              Ratio:  2.447                Gaps:      7
 Percent Similarity:  59.119   Percent Identity: 37.107

alignment_block:
 US-09-215-435-225 x AC006915   ..

 Align seg 1/1  to: AC006915  from: 1  to: 134095

      49 PheTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAs 65
         |||      :::||||||||||:::        ::::::|||
   99439 TTTCAGGCTAATCTTGGAAACGTT.........CTCACTCCAACC..... 99474

      65 nAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPheP 82
            :::   |||   |||           |||   ||||||:::
   99475 .......CAAGTAAAGGATACT...........CCAGAAGTAAAATGG. 99504

      82 roGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAsp 98
           |||      ||||||   |||   |||:::   :::||||||||||
   99505 ..GATGCTGAGCCAGGAGCACTCTACACTCTCATCAAGACGGATCCAGAT 99552

      99 AlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuVa 115
         ||||||||||||   ||||||      |||   |||:::||||||||||||
   99553 GCTCCATCTCGCAAAGAGCCAACCTACCGAGAGTGGCATCACTGGCTCGT 99602

     115 lThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnG 132
         |:::::|||   |||   |||:::   |||             |||:::
   99603 CGTCAACATTCCAGGAAACGACATTGCCAAA...........GGAGACA 99640

     132 luLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHis 148
           ||||||   |||   :::   :::||||||      :::|||:::|||
   99641 CTCTTTCCGAGTACATCGGCGCCGGACCACCACCAAAAACTGGACTCCAT 99690

     149 ArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer..... 163
         ||||||   ::::::::::|||   |||:::|||::::::   :::
   99691 CGTTACGTTTACCTGATCTACAAGCAATCTGGACGCATCGAAGATGCCGA 99740

     164 .........LeuLeuProLysGluAsnLysThrArgGlySerTrpLysM 177
                  |||         ::::::      ||||||:::||||||
   99741 GCACGGACGTCTAACCAACACTTCCGGAGACAAAAGAGGAGGATGGAAGG 99790

     177 etAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThr 193
               |||:::   :::      ||||||   |||      ::::::
   99791 CTGCCGATTTCGTCGCAAAGCACAAGCTCGGTGCTCCAGTCTTCGGAAAT 99840

     194 GlnPheMetThrGlnAsnTyrGlnAsp 202
         |||::::::::::   |||:::|||
   99841 CTGTTCCAGGCCGAG...TACGACGAC 99864

seq_name: gb_htg4:AC012978

seq_documentation_block:
LOCUS       AC012978    63551 bp    DNA             HTG       03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC012978
```

```
VERSION     AC012978.1  GI:6223345
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 63551)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212016 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .63551
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    18144 a  13546 c  14099 g  17762 t
ORIGIN


alignment_scores:
            Quality:  229.00              Length:    131
              Ratio:  2.516                 Gaps:      5
 Percent Similarity:  69.466   Percent Identity: 40.458

alignment_block:
 US-09-215-435-225 x AC012978/rev   ..

 Align seg 1/1  to reverse of: AC012978  from: 1  to: 63551

      76 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 92
         ||||||   |||||||:::      |||      :::      |||   ||
   23000 GAGCCGTGCGTGAAATGGGAG...GCGGACGCCAATAAGCTGTACACCCT 22954

      92 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 109
         |   |||:::|||||||||||||||||||||||   :::|||:::   |||
   22953 CTGCATGACCGATCCGGATGCGCCCAGTCGCAAGGATCCCAAGTTTAGGG 22904

     109 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 125
         |||:::|||||||||||||   :::|||   |||:::|||:::   |||
   22903 AGTGGCACCATTGGCTGGTGGGCAACATACCCGGTGGAGATGTCGCCAAG 22854

     126 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 142
         |||::::::                |||||||||||   :::   :::|||||
   22853 GGCGAGGTT............CTCTCCGCCTACGTGGGATCCGGGCCTCC 22816

     142 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 159
         |      :::|||:::|||||||||   |||:::::::|||   |||:::
   22815 ACCAGACACCGGACTCCATCGTTACGTTTTCCTGATCTACGAGCAGCGGT 22766

     159 lyLysValIle.........SerLeuLeuProLysGluAsn...LysThr 171
         |||:::            :::   ||||||:::::::::
   22765 GCAAGCTCACATTCGACGAGAAGCGACTGCCCAATAACAGCGGAGATGGA 22716

     172 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 188
         ||||||::::::|||:::   :::|||   :::::::::   |||||||::
   22715 CGCGGTGGCTTCAAAATCGCCGAGTTCGCCAAGAAGTACGCCCTCGGCAA 22666

     188 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAsp 202
         :|||   |||::::::   ::::::::::::   |||:::|||
   22665 TCCCATTGCCGGTAACCTCTACCAGGCTGAG...TACGACGAC 22626

 seq_name: gb_in1:AC005891

 seq_documentation_block:
 LOCUS       AC005891   136933 bp    DNA             INV       30-OCT-1998
```

```
DEFINITION  Drosophila melanogaster, chromosome 2L, region 33F3-34A2, P1 clones
            DS08479 and DS07153, complete sequence.
ACCESSION   AC005891 AC005108 AC005109 AC005110
VERSION     AC005891.1  GI:3810575
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 136933)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2L, region 33F3-34A2
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 136933)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library locations: 89-31, 75-49.
FEATURES             Location/Qualifiers
     source          1. .136933
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="33F3-34A2"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
                     /clone="P1s DS08479 (D303) and DS07153 (D305)"
                     /note="DS08479 (D303) and DS07153 (D305) were completed as
                     a project.  DS08479 extends from P1 end at bp 1 to P1 end
                     at bp 83797.  DS07153  extends from p1 end at bp 51,253 to
                     P1 end at bp 136,933."
BASE COUNT    37380 a  28396 c  29033 g  42124 t
ORIGIN


alignment_scores:
            Quality:  229.00              Length:    131
              Ratio:  2.516                 Gaps:      5
 Percent Similarity:  69.466   Percent Identity: 40.458

alignment_block:
 US-09-215-435-225 x AC005891   ..

 Align seg 1/1  to: AC005891  from: 1  to: 136933

      76 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 92
```

```
      ||||||   ||||||:::      |||      :::      |||   ||
24363 GAGCCGTGCGTGAAATGGGAG...GCGGACGCCAATAAGCTGTACACCCT 24409

   92 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 109
      |   |||:::|||||||||||||||||||||||   :::|||:::   |||
24410 CTGCATGACCGATCCGGATGCGCCCAGTCGCAAGGATCCCAAGTTTAGGG 24459

  109 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 125
        |||:::||||||||||||   :::|||   |||:::|||:::   |||
24460 AGTGGCACCATTGGCTGGTGGGCAACATACCCGGTGGAGATGTCGCCAAG 24509

  126 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 142
      |||::::::            ||||||||||   :::   :::|||||
24510 GGCGAGGTT............CTCTCCGCCTACGTGGGATCCGGGCCTCC 24547

  142 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 159
      |      :::|||:::|||||||||   |||:::::|||   |||:::
24548 ACCAGACACCGGACTCCATCGTTACGTTTTCCTGATCTACGAGCAGCGGT 24597

  159 lyLysValIle.........SerLeuLeuProLysGluAsn...LysThr 171
        |||:::            :::   ||||||:::::::::
24598 GCAAGCTCACATTCGACGAGAAGCGACTGCCCAATAACAGCGGAGATGGA 24647

  172 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 188
      ||||||:::::|||:::   :::|||   :::::::::   |||||||::
24648 CGCGGTGGCTTCAAAATCGCCGAGTTCGCCAAGAAGTACGCCCTCGGCAA 24697

  188 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAsp 202
      :|||   |||::::::   ::::::::::::   |||:::|||
24698 TCCCATTGCCGGTAACCTCTACCAGGCTGAG...TACGACGAC 24737
```

THIS PAGE BLANK (USPTO)

OM of: US-09-215-435-225 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:29 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480	+	1224.00	2543.16	1.1e-133	681	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660	+	1224.00	2541.04	1.4e-133	826	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013	+	1224.00	2541.04	1.4e-133	826	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252	+	1224.00	2541.04	1.4e-133	826	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490	+	1224.00	2540.14	1.6e-133	897	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882	+	1220.00	2531.34	5.0e-133	932	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426	+	1218.00	2526.07	9.9e-133	1028				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539	+	1215.00	2521.14	1.9e-132	909	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582	+	1213.00	2517.02	3.1e-132	903	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431	+	1198.00	2485.01	1.9e-130	952	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350	+	1189.00	2466.89	2.0e-129	890	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582	+	652.00	1349.31	3.5e-67	447	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30001	+	237.50	471.50	2.7e-18	1047				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58685	+	237.50	471.50	2.7e-18	1047				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30002	+	208.50	407.19	1.0e-14	1447				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58686	+	208.50	407.19	1.0e-14	1447				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80734	+	207.50	409.14	8.1e-15	1001				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ14832	+	201.50	398.73	3.1e-14	822	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60684	+	182.00	362.72	3.1e-12	528	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC42966	+	182.00	362.72	3.1e-12	528	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60683	+	164.00	325.01	3.9e-10	528	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC48359	+	164.00	321.10	6.5e-10	754	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34682	+	164.00	320.80	6.7e-10	775	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28490	+	164.00	319.71	7.7e-10	856	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28491	-	164.00	319.71	7.7e-10	856	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT60139	+	148.00	285.29	6.4e-08	929	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90668	+	147.50	290.63	3.2e-08	519	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90636	+	145.50	286.37	5.6e-08	522	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT60140	+	140.50	273.19	3.0e-07	668	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC93828	+	121.00	238.84	2.5e-05	369	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA87678	+	90.50	162.01	0.4713	1200				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX39569	+	90.00	173.83	0.1034	371	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV63925	+	89.50	165.06	0.3183	750	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX81032	+	89.50	165.06	0.3183	750	!			
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85554	+	89.50	158.41	0.7475	1376				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA87700	+	88.00	157.15	0.8787	1159				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ01425	-	88.00	82.59	1.2e+04	10386				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF27717	+	87.50	142.45	5.79	4023				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ51531	+	87.50	140.53	7.41	4795				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX16969	+	87.50	140.53	7.41	4795				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC53434	+	86.00	152.20	1.66	124				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77512	+	85.00	136.86	11.86	415				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA02705	+	83.50	153.16	1.47	706				
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT34817	+	83.50	152.55	1.58	746				
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC78058	+	83.50	143.90	4.81	164				

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480

seq_documentation_block:
ID AAF94480 standard; cDNA; 681 BP.
XX
AC AAF94480;
XX
DT 04-JUN-2001 (first entry)
XX
DE Human hydrophobic domain containing protein clone HP03880 cDNA #104.
XX
KW Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW behavioural characteristic; immune response; ss.
XX
OS Homo sapiens.
XX
PN WO200112660-A2.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-JP05356.
XX
PR 17-AUG-1999; 99JP-0230344.
PR 07-SEP-1999; 99JP-0252551.
PR 01-OCT-1999; 99JP-0281132.
PR 22-OCT-1999; 99JP-0301624.
PR 04-NOV-1999; 99JP-0313877.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2001-160059/16.
DR P-PSDB; AAB88590.
XX
PT Human proteins with hydrophobic domains and the DNAs which encode them
PT are useful for treating autoimmune disorders, burns and tumors and for
PT screening novel pharmaceuticals -
XX
PS Claim 3; Page 426-427; 518pp; English.
XX
CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC osteopathic, anti-inflammatory and cytostatic activities, and can be
CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC probes for genetic diagnosis and gene sources for gene therapy or for
CC producing (I) in large quantities. Cells containing (II) are used for
CC the detection of ligands or receptors corresponding to membrane or
CC secretory proteins and to screen small molecule novel pharmaceuticals.
CC Antibodies directed to (I) can be used for the detection, quantification
CC and purification of (I). Activities of (I) may include cytokine and cell
CC proliferation/differentiation function, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC wound healing, as nutritional sources or supplements e.g. as amino acid,
CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,

```
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 681 BP; 179 A; 186 C; 183 G; 133 T; 0 other;


alignment_scores:
           Quality: 1224.00              Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AAF94480   ..

 Align seg 1/1  to: AAF94480  from: 1  to: 681

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 50

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 100

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 150

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 200

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 250

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 300

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 350

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 400

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 450

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 500

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 550

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 600

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 650

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     651 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 681
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660

seq_documentation_block:
ID   AAX97660 standard; DNA; 826 BP.
XX
AC   AAX97660;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 124.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY35976.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 255; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
           Quality: 1224.00              Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AAX97660   ..

 Align seg 1/1  to: AAX97660  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64
```

```
     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013

seq_documentation_block:
ID   AAC00013 standard; cDNA; 826 BP.
XX
AC   AAC00013;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein cDNA sequence #4.
XX
KW   Human; secreted protein; 5' EST; expressed sequence tag; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             15..698
FT                   /*tag= a
FT                   /product= "secreted protein"
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00016.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Example 19; SEQ ID 7; 71pp + CD-ROM; English.
XX
CC   The present sequence is a full length cDNA encoding a human
CC   secreted protein. The cDNA was obtained from a 5' EST using first and
CC   second strand synthesis procedures. 5' ESTs were prepared from total
CC   human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region
CC   (UTR) of the mRNA because they are often obtained from oligo-dT primed
CC   cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC   sequences derived from the 5' ends of mRNAs and even in those cases where
CC   longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC   included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC   therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC   are also used in diagnostic, forensic, gene therapy and chromosome
CC   mapping procedures. They are used to obtain upstream regulatory sequences
CC   and to design expression and secretion vectors.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AAC00013   ..

 Align seg 1/1  to: AAC00013  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314
```

```
    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252

seq_documentation_block:
ID   AAZ42252 standard; cDNA; 826 BP.
XX
AC   AAZ42252;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human phosphatidylethanolamine-binding protein encoding cDNA.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   P-PSDB; AAY64647.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Example 21; Page 168-169; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AAZ42252   ..

 Align seg 1/1  to: AAZ42252  from: 1  to: 826

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

      184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

      201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

      217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
          |||||||||||||||||||||||||||||||
      665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490

seq_documentation_block:
ID   AAF94490 standard; cDNA; 897 BP.
XX
AC   AAF94490;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Human hydrophobic domain containing protein clone HP03880 cDNA #114.
XX
KW   Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW   antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW   cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW   HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW   inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW   behavioural characteristic; immune response; ss.
XX
OS   Homo sapiens.
XX
PN   WO200112660-A2.
XX
PD   22-FEB-2001.
XX
PF   10-AUG-2000; 2000WO-JP05356.
XX
PR   17-AUG-1999;   99JP-0230344.
PR   07-SEP-1999;   99JP-0252551.
PR   01-OCT-1999;   99JP-0281132.
PR   22-OCT-1999;   99JP-0301624.
PR   04-NOV-1999;   99JP-0313877.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-160059/16.
DR   P-PSDB; AAB88590.
XX
PT   Human proteins with hydrophobic domains and the DNAs which encode them
PT   are useful for treating autoimmune disorders, burns and tumors and for
PT   screening novel pharmaceuticals -
XX
PS   Claim 4; Page 442-444; 518pp; English.
XX
CC   AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC   AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC   anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC   osteopathic, anti-inflammatory and cytostatic activities, and can be
CC   used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC   to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC   probes for genetic diagnosis and gene sources for gene therapy or for
CC   producing (I) in large quantities. Cells containing (II) are used for
CC   the detection of ligands or receptors corresponding to membrane or
CC   secretory proteins and to screen small molecule novel pharmaceuticals.
CC   Antibodies directed to (I) can be used for the detection, quantification
CC   and purification of (I). Activities of (I) may include cytokine and cell
CC   proliferation/differentiation function, immune stimulating or suppressing
CC   activity, haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC   activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC   multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC   inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC   wound healing, as nutritional sources or supplements e.g. as amino acid,
CC   carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 897 BP; 226 A; 253 C; 234 G; 184 T; 0 other;

alignment_scores:
           Quality: 1224.00              Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x AAF94490   ..

 Align seg 1/1  to: AAF94490  from: 1  to: 897

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 148

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     149 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 198

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     199 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 248

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     249 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 298

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     299 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 348

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     349 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 398

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     399 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 448

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     449 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 498

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     499 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 548

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     549 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 598

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     599 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 648

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     649 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 698

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     699 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 748

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     749 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 779

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882

seq_documentation_block:
ID   AAC84882 standard; cDNA; 932 BP.
XX
AC   AAC84882;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human SEC1 nucleic acid sequence (clone ID 3445452).
XX
KW   SECX; cytostatic; gynecological; gene therapy; screening assay; human;
KW   chromosomal mapping; forensic biology; cell proliferation; cancer;
KW   cell differentiation; immune associated disorder; gestational disease;
KW   SEC1; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             113..781
FT                   /*tag=  a
FT                   /product= "SEC1"
FT   sig_peptide     113..181
FT                   /*tag=  b
FT   mat_peptide     182..778
FT                   /*tag=  c
XX
PN   WO200078802-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17328.
XX
PR   23-JUN-1999;   99US-0140584.
PR   20-JUL-1999;   99US-0144722.
PR   16-SEP-1999;   99US-0154520.
PR   22-JUN-2000; 2000US-0604286.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Vernet C,  Yang M,  Boldog FL;
PI   Herrmann JL;
XX
DR   WPI; 2001-071385/08.
DR   P-PSDB; AAB48368.
XX
PT   Polynucleotides encoding SECX proteins useful for treating disease
PT   characterized by an aberrant level of cell proliferation and/or
PT   differentiation like cancer or immune associated disorders -
XX
PS   Claim 3; Fig 1; 132pp; English.
XX
CC   The invention relates to human SECX polypeptides and polynucleotides
CC   encoding them. The SECX polypeptides can be expressed by standard
CC   recombinant methodology.  The SECX polypeptides are useful for treating
CC   or preventing a SECX-associated disorder. The invention is useful in
CC   screening assays; detection assays (e.g. chromosomal mapping, cell and
CC   tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC   prognostic assays, monitoring clinical trials, and pharmocogenomics); and
CC   methods of treatment (e.g. therapeutic and prophylactic), especially
CC   disorders characterized by aberrant cell proliferation and/ or
CC   differentiation like cancer or immune associated disorders or gestational
CC   disease. The present sequence represents a SEC1 nucleic acid sequence.
XX
SQ   Sequence 932 BP; 244 A; 257 C; 244 G; 187 T; 0 other;


alignment_scores:
            Quality: 1220.00              Length:    227
              Ratio:  5.374                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x AAC84882    ..

 Align seg 1/1  to: AAC84882  from: 1  to: 932

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     113 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 162

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     163 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 212

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     213 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 262

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     263 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 312

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     313 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 362

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     363 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 412

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     413 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 462

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||:::||||||||||||||||||||||||||
     463 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 512

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 562

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     563 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 612

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     613 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 662

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     663 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 712

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     713 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 762

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     763 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 793

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426

 seq_documentation_block:
```

```
ID   AAA78426 standard; cDNA; 1028 BP.
XX
AC   AAA78426;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:56.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24482.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 475; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1028 BP; 309 A; 268 C; 259 G; 186 T; 6 other;

alignment_scores:
            Quality: 1218.00              Length:    227
              Ratio:  5.389                Gaps:      0
 Percent Similarity:  99.559    Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x AAA78426   ..

 Align seg 1/1  to: AAA78426  from: 1  to: 1028

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          153 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 202

                                                           17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          203 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 252

                                                           34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          253 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 302

                                                           51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          303 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 352

                                                           67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          353 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 402

                                                           84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          403 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 452

                                                          101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          453 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 502

                                                          117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
                                                              ||||||||||||||||||||    ||||||||||||||||||||||||||||
                                                          503 TATCAAGGGCGCCGACCTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTAT 552

                                                          134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          553 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 602

                                                          151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          603 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 652

                                                          167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          653 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 702

                                                          184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          703 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 752

                                                          201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                          753 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 802

                                                          217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
                                                              ||||||||||||||||||||||||||||||
                                                          803 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 833

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539

seq_documentation_block:
ID   AAC77539 standard; cDNA; 909 BP.
XX
AC   AAC77539;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF3094 polynucleotide sequence SEQ ID NO:6187.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
```

```
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43330.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5369-5370; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 909 BP; 232 A; 254 C; 236 G; 186 T; 1 other;


alignment_scores:
            Quality: 1215.00              Length:    227
              Ratio:  5.376                Gaps:      0
 Percent Similarity:  99.559   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x AAC77539   ..

 Align seg 1/1  to: AAC77539  from: 1  to: 909

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     100 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 149

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     150 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 199

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     200 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 249

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     250 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 299

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     300 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 349

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         |||||   ||||||||||||||||||||||||||||||||||||||||||
     350 CCGTGTACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 399

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     400 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 449

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     450 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 499

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     500 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 549

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     550 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 599

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     600 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 649

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     650 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 699

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     700 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 749

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     750 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 780

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582

seq_documentation_block:
ID   AAA15582 standard; cDNA; 903 BP.
XX
AC   AAA15582;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human phospholipid binding protein 2, PLBP2 gene.
XX
KW   Human; phospholipid binding protein; PLBP2; foetal development disorder;
KW   reproduction disorder; cell proliferation disorder; immune response;
KW   autoimmune disorder; AIDS; infertility; cytostatic; immunosuppressive;
KW   gene therapy; hereditary neuropathy;
KW   phosphatidylethanolamine binding protein D1; PE-BP D1; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..771
```

```
FT                  /*tag=  a
FT                  /product= "Human PLBP2"
XX
PN   US6063767-A.
XX
PD   16-MAY-2000.
XX
PF   09-DEC-1998;   98US-0208718.
XX
PR   28-OCT-1997;   97US-0958820.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Corley NC,  Shah P,  Lal P,  Hillman JL;
XX
DR   WPI; 2000-375529/32.
DR   P-PSDB; AAY94263.
XX
PT   New purified phospholipid binding proteins 1 and 2 useful for
PT   diagnosing, treating or preventing diseases disorders associated with
PT   fetal development, reproduction, cell proliferation, and the immune
PT   response  -
XX
PS   Example 5; Fig 2; 37pp; English.
XX
CC   The present sequence is the phospholipid binding protein 2 (PLBP2) gene.
CC   This gene is expressed in lung, prostate and heart tissues. Also, the
CC   protein is expressed in foetal tumour tissues. PLBP2 may be used for the
CC   diagnosis, prevention, or treatment of disorders associated with foetal
CC   development (e.g. hereditary neuropathies), reproduction (e.g.
CC   infertility), cell proliferation (e.g. cancers), and the immune response
CC   (AIDS). PLBP2 antibodies may also be developed for potential drug
CC   screening or to quantitate PLBP2 gene expression in biopsied tissues. The
CC   PLBP2 gene may be administered for gene therapy of disorders associated
CC   with PLBP2. PLBP2 has high homology with the phosphatidylethanolamine
CC   binding protein D1, PE-BP D1, of Onchocerca volvulus. PE-BP DI is thought
CC   to play a role in transport or signal mechanisms between membranes and
CC   the cytoplasm.
XX
SQ   Sequence 903 BP; 222 A; 251 C; 242 G; 188 T; 0 other;


alignment_scores:
            Quality: 1213.00              Length:    227
              Ratio:  5.367                Gaps:      0
 Percent Similarity:  99.559   Percent Identity: 99.119

alignment_block:
 US-09-215-435-225 x AAA15582   ..

 Align seg 1/1  to: AAA15582  from: 1  to: 903

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 137

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 187

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 237

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     238 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 287

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     288 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 337

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     338 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 387

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     388 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 437

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         |||||||||||||||||||||:::|||||||||||||||||||||||||
     438 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 487

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 537

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     538 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 587

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     588 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCT 637

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     638 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 687

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||  ||||||||||||||||
     688 CAGGACTCACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAA 737

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||:||||||||||||
     738 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 768

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431

seq_documentation_block:
ID   AAA78431 standard; cDNA; 952 BP.
XX
AC   AAA78431;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:61.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24487.
XX
```

```
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 478; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 952 BP; 310 A; 247 C; 223 G; 172 T; 0 other;


alignment_scores:
            Quality: 1198.00              Length:    229
              Ratio:  5.278                Gaps:      2
 Percent Similarity:  99.127   Percent Identity: 99.127

alignment_block:
 US-09-215-435-225 x AAA78431   ..

 Align seg 1/1  to: AAA78431  from: 1  to: 952

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      63 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 112

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     113 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 162

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     163 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 212

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     213 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 262

      67 rArgGlnLysIleThrSerTrpMetGlu.ProIleValLysPheProGly 83
         |||||||||||||||||||||||||| ||||||||||||||||||||||
     263 CAGACAGAAGATCACCTCCTGGATGGAAGCCGATAGTCAAGTTCCCGGGG 312

      84 AlaValAspGlyAlaThrTyr.IleLeuValMetValAspProAspAlaP 100
         |||||||||||||||||||| ||||||||||||||||||||||||||||
     313 GCCGTGGACGGCGCAACCTATAATCCTGGTGATGGTGGATCCAGATGCCC 362

     100 roSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThr 116
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     363 CTAGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACA 412

     117 AspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLe 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     413 GATATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTT 462

     133 uSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgT 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     463 ATCAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCT 512

     150 yrGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuPro 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 ACCAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCC 562

     167 LysGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnAr 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     563 AAGGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCG 612

     183 gPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     613 TTTCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACT 662

     200 yrGlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluPro 216
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     663 ACCAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCC 712

     217 LysHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||||
     713 AAGCACAAAAACCAGGCGGAGATAGCTGCCTGC 745

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350

seq_documentation_block:
ID   AAA96350 standard; cDNA; 890 BP.
XX
AC   AAA96350;
XX
DT   08-FEB-2001  (first entry)
XX
DE   cDNA encoding a novel polypeptide designated PRO4408.
XX
KW   Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW   PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW   PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW   PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW   insulinemia; kidney disorder; Bergers disease; nephropathy;
KW   Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW   Crohns disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             89..760
FT                   /*tag=  a
FT   sig_peptide     89..154
FT                   /*tag=  b
XX
PN   WO200056889-A2.
XX
PD   28-SEP-2000.
XX
PF   01-MAR-2000; 2000WO-US05601.
XX
PR   23-MAR-1999;   99US-0125774.
PR   23-MAR-1999;   99US-0125778.
PR   24-MAR-1999;   99US-0125826.
PR   31-MAR-1999;   99US-0127035.
PR   05-APR-1999;   99US-0127706.
PR   21-APR-1999;   99US-0130359.
PR   27-APR-1999;   99US-0131270.
PR   27-APR-1999;   99US-0131272.
PR   27-APR-1999;   99US-0131291.
PR   04-MAY-1999;   99US-0132371.
PR   04-MAY-1999;   99US-0132379.
PR   04-MAY-1999;   99US-0132383.
PR   25-MAY-1999;   99US-0135750.
PR   08-JUN-1999;   99US-0138166.
PR   20-JUL-1999;   99US-0144791.
PR   03-AUG-1999;   99US-0146970.
PR   09-DEC-1999;   99US-0170262.
XX
PA   (GETH ) GENENTECH INC.
XX
```

```
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2000-628263/60.
DR   P-PSDB; AAB18923.
XX
PT   Novel secreted and transmembrane polypeptides useful for diagnosing
PT   tumour in a mammal, for identifying agonists and antagonists of the
PT   polypeptide and for therapeutic use  -
XX
PS   Claim 2; Fig 29; 222pp; English.
XX
CC   The present sequence encodes a secreted or transmembrane polypeptide.
CC   The specification describes polypeptides designated PRO1484, PRO4334,
CC   PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC   PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC   PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC   useful for diagnosing tumour in a mammal. The polypeptides, their
CC   agonists and antagonists are useful treating a condition associated with
CC   expression or activity of the polypeptide. Conditions treated include
CC   obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC   capable of inducing proliferation of mammalian kidney mesangial cells
CC   and are therefore useful for treating kidney disorders associated with
CC   decreased mesangial cell function such as Bergers disease or other
CC   nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC   dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC   to generate transgenic animals for use in development and screening of
CC   therapeutically useful reagents and also for chromosome identification
CC   and tissue typing.
XX
SQ   Sequence 890 BP; 228 A; 246 C; 234 G; 182 T; 0 other;


alignment_scores:
            Quality: 1189.00              Length:    227
              Ratio:  5.261                Gaps:      0
 Percent Similarity:  99.559    Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x AAA96350   ..

 Align seg 1/1  to: AAA96350  from: 1  to: 890

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      89 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 138

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     139 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 188

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     189 CCCTCTTGGACGAGGACACCCTCTTTGCCAGGGCCTTGAAGTTTTCTAC 238

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     239 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 288

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     289 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 338

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     339 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 388

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     389 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 438

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     439 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 488

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     489 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 538

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     539 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 588

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     589 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCT 638

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     639 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 688

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||  ||||||||||||||||
     689 CAGGACTCACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAA 738

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         ||||   |||||||||||||||||||||||||
     739 GCACAA.AACCAGGCAGAGATAGCTGCCTGC 768

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582

seq_documentation_block:
ID   AAX40582 standard; cDNA; 447 BP.
XX
AC   AAX40582;
XX
DT   18-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID No: 182.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906550-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01232.
XX
PR   01-AUG-1997;   97US-0905144.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153780/13.
DR   P-PSDB; AAY11860.
XX
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
XX
PS   Claim 1; Page 285; 675pp; English.
XX
CC   AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins expressed in prostate, and encode the proteins
CC   given in AAY11716 to AAY11993 respectively. The proteins given represent
CC   the signal peptide and an N-terminal fragment of a secreted protein. The
CC   nucleic acid sequences can be used for producing secreted human gene
CC   products. They can also be used to develop products for diagnosis and
```

```
CC   therapy. The proteins obtained may have cytokine activity, cell
CC   proliferation and differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptides can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 447 BP; 98 A; 110 C; 136 G; 99 T; 4 other;


alignment_scores:
            Quality:  652.00              Length:    121
              Ratio:  5.433                Gaps:      0
 Percent Similarity:  99.174   Percent Identity: 99.174

alignment_block:
 US-09-215-435-225 x AAX40582   ..

 Align seg 1/1  to: AAX40582  from: 1  to: 447

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 134

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     135 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 184

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 234

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||   |||||
     235 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTDACAACTA 284

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     285 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 334

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     335 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 384

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     385 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 434

     117 pIleLysGlyAla 121
         ||||||||||||
     435 TATCAAGGGCGCC 447

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30001

seq_documentation_block:
ID   AAQ30001 standard; cDNA to mRNA; 1047 BP.
XX
AC   AAQ30001;
XX
DT   18-MAR-1993  (first entry)
XX
DE   HCNP precursor gene #1.
XX
KW   Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;
KW   neonatal rat; expression vector; neurodegenerative disorder; dementia;
KW   Alzheimer's disease; Parkinson's disease; ss.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             26..589
FT                   /*tag= a
FT   5'UTR           1..25
FT                   /*tag= b
FT   3'UTR           587..1047
FT                   /*tag= c
FT   polyA_signal    987..992
FT                   /*tag= d
FT   misc_RNA        29..61
FT                   /*tag= e
FT                   /label= Neurotrophic_activity_region
XX
PN   EP511816-A.
XX
PD   04-NOV-1992.
XX
PF   27-APR-1992;   92EP-0303800.
XX
PR   27-APR-1991;   91JP-0124688.
XX
PA   (OJIK/) OJIKA K.
PA   (SUMU ) SUMITOMO PHARM CO LTD.
PA   (YAMA/) YAMAMOTO M.
XX
PI   Agui H,  Fukushima N,  Irie T,  Kojima S,  Nishihara T;
PI   Ojika K,  Ono K,  Tohdoh N,  Tojo S,  Ueki Y;
XX
DR   WPI; 1992-367633/45.
DR   P-PSDB; AAR27897.
XX
PT   Neurotrophic peptide derivs. - used for treating neurological
PT   degenerative disorders, e.g. Alzheimer's disease or Parkinson's
PT   disease
XX
PS   Claim 3; Page 31-32; 57pp; English.
XX
CC   The seqence given is the hippocampal cholinergic neurotrophic peptide
CC   (HCNP) gene.  The active part of the protein encoded by this sequence
CC   is located in the first eleven amino acids at the N terminus.  This
CC   gene was obtained by preparing mRNA from the hippocampal tissue of
CC   neonatal rats, 12 days after birth.  The mRNA is converted to cDNA by
CC   standard methods.  The cDNA can be used in expression vectors to
CC   produce both the full length protein and the active peptide fragment.
CC   The gene and the peptide are useful for the treatment of
CC   neurodegenerative disorders and dementia eg. Alzheimer's and
CC   Parkinson's disease.
XX
SQ   Sequence 1047 BP; 257 A; 296 C; 290 G; 204 T; 0 other;


alignment_scores:
            Quality:  237.50              Length:    185
              Ratio:  2.284                Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225 x AAQ30001   ..

 Align seg 1/1  to: AAQ30001  from: 1  to: 1047

      52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
         |||||||||::::::         :::::::|||               ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
         |   :::      ||||||                          |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
            |||      |||   ||||||::::::|||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250
```

```
     102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

     118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
         :||||||   |||::::::::::|||   :::           ||||||
     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC...........CTCTCCG 338

     135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 151
           |||   :::   :::|||||||      :::|||:::|||||||||
     339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

     152 PhePheValTyrLeuGlnGluGlyLysValIleSer...........Le 164
         ::::::||||||   ||||||      :::                 ||
     389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

     164 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 181
         |      |||::::::   :::||||||::::::|||::::::   |||
     439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

     181 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 197
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

     198 GlnAsnTyrGlnAspSer.............................. 203
         :::   ::::::||||||
     539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585

     204 ..............................ProThrLeuGlnAlaP 209
                                      ||||||:::|||:::
     586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

     209 roArg 210
           |||
     636 GTCGT 640
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58685

seq_documentation_block:

```
ID   AAQ58685 standard; cDNA to mRNA; 1047 BP.
XX
AC   AAQ58685;
XX
DT   13-OCT-1994  (first entry)
XX
DE   Rat hippocampal cholinergic neurotrophic peptide precursor cDNA.
XX
KW   Rat hippocampal cholinergic neurotrophic peptide; rat HCNP;
KW   nerve degeneration; acetylcholine synthesis; neurostimulation; ds.
XX
OS   Rattus norvegicus (Wistar).
XX
FH   Key             Location/Qualifiers
FT   CDS             26..589
FT                   /*tag=  a
FT                   /product= rat_HCNP
FT   5'UTR           1..25
FT                   /*tag=  b
FT   3'UTR           590..1047
FT                   /*tag=  c
FT   polyA_signal    987..992
FT                   /*tag=  d
FT   polyA_site      1008..1047
FT                   /*tag=  e
XX
PN   WO9405788-A.
XX
PD   17-MAR-1994.
XX
PF   27-AUG-1993;   93WO-JP01214.
XX
PR   29-AUG-1992;   92JP-0253734.
PR   28-AUG-1992;   92JP-0254170.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
PA   (YAMA/) YAMAMOTO M.
XX
PI   Agui H,  Fukushima N,  Irie T,  Kamikawa Y,  Kojima S;
PI   Nishihara T,  Ojika K,  Ono K,  Taiji M,  Tohdoh N;
PI   Tojo S,  Ueki Y;
XX
DR   WPI; 1994-101193/12.
DR   P-PSDB; AAR49942.
XX
PT   Neurotrophic peptide(s), precursors and genes - used to treat
PT   nervous degeneration, increases acetylcholine synthesis
XX
PS   Claim 3; Page 153-154; 200pp; Japanese.
XX
CC   The first 11 N-terminal amino acids of the rat hippocampal
CC   cholinergic neurotrophic peptide precursor (encoded by AAQ58685)
CC   have neurotrophic activity and increase acetylcholine synthesis.
XX
SQ   Sequence 1047 BP; 257 A; 297 C; 289 G; 204 T; 0 other;
```

alignment_scores:

```
            Quality:  237.50              Length:    185
              Ratio:  2.284                 Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973
```

alignment_block:

```
 US-09-215-435-225 x AAQ58685   ..

 Align seg 1/1  to: AAQ58685  from: 1  to: 1047

      52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
         |||||||||::::::         ::::::|||                ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
         |   :::      ||||||                              |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
           |||      |||   ||||||:::::||||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

     102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

     118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
         :||||||   |||::::::::::|||   :::           ||||||
     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC...........CTCTCCG 338

     135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 151
           |||   :::   :::|||||||      :::|||:::|||||||||
     339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

     152 PhePheValTyrLeuGlnGluGlyLysValIleSer...........Le 164
         ::::::||||||   ||||||      :::                 ||
     389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

     164 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 181
         |      |||::::::   :::||||||::::::|||::::::   |||
     439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

     181 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 197
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

     198 GlnAsnTyrGlnAspSer.............................. 203
         :::   ::::::||||||
     539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585
```

```
     204 ..............................ProThrLeuGlnAlaP 209
                                       ||||||:::|||:::
     586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

     209 roArg 210
           |||
     636 GTCGT 640

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30002

seq_documentation_block:
ID   AAQ30002 standard; cDNA to mRNA; 1447 BP.
XX
AC   AAQ30002;
XX
DT   18-MAR-1993  (first entry)
XX
DE   HCNP precursor gene #2.
XX
KW   Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;
KW   expression vector; neurodegenerative disorder; dementia;
KW   Alzheimer's disease; Parkinson's disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             120..680
FT                   /*tag= a
FT   5'UTR           1..119
FT                   /*tag= b
FT   3'UTR           681..1447
FT                   /*tag= c
FT   misc_RNA        123..155
FT                   /*tag= d
FT                   /label= HCNP
XX
PN   EP511816-A.
XX
PD   04-NOV-1992.
XX
PF   27-APR-1992;   92EP-0303800.
XX
PR   27-APR-1991;   91JP-0124688.
XX
PA   (OJIK/) OJIKA K.
PA   (SUMU ) SUMITOMO PHARM CO LTD.
PA   (YAMA/) YAMAMOTO M.
XX
PI   Agui H,  Fukushima N,  Irie T,  Kojima S,  Nishihara T;
PI   Ojika K,  Ono K,  Tohdoh N,  Tojo S,  Ueki Y;
XX
DR   WPI; 1992-367633/45.
DR   P-PSDB; AAR27718.
XX
PT   Neurotrophic peptide derivs. - used for treating neurological
PT   degenerative disorders, e.g. Alzheimer's disease or Parkinson's
PT   disease
XX
PS   Disclosure; Page 42-44; 57pp; English.
XX
CC   The sequence given is the human hippocampal cholinergic neurotrophic
CC   peptide (HCNP) precursor gene.  The active part of the protein encoded
CC   by this sequence is located in the first eleven amino acids at the N
CC   terminus.  This gene was isolated using the rat HCNP precusor gene
CC   sequence as a guide (see also AAQ30001).  From studies of the
CC   structure of the rat HCPN precusor gene and protein the human HCNP
CC   was determined.  The rat and human HCNP's have markedly different
CC   structures and composition.  HCNP's are useful for the treatment of
CC   neurodegenerative disorders and dementia eg. Alzheimer's and
CC   Parkinson's disease.
XX
SQ   Sequence 1447 BP; 342 A; 351 C; 406 G; 348 T; 0 other;

alignment_scores:
            Quality:  208.50              Length:    206
              Ratio:  1.813                 Gaps:      7
 Percent Similarity:  55.825   Percent Identity: 31.553

alignment_block:
 US-09-215-435-225 x AAQ30002   ..

 Align seg 1/1  to: AAQ30002  from: 1  to: 1447

       3 TrpThrMetArgLeuValThrAlaAla.LeuLeuLeuGlyLeuMetMetV 19
         |||             :::   ||||||   ||||||||||||   |||
      74 TGGCCTACCGCGGCACTCCCGGCTGCACGCTCTGCTTGGCCTCGCCATGC 123

      19 alValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeu 35
         |||          |||   ::::::     :::      :::|||
     124 CGGTG.........GACCTCAGCAAGTGGTCCGGGCCCTTGAGCCTG 161

      36 LeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr..... 50
         :::   |||           :::   |||:::|||   |||
     162 CAAGAAGTGGACGAGCAGCCGCAGCACCCGCTGCATGTCACCTACGCCGG 211

      51 .........ProGluLeuGlyAsnIleGlyCysLysValValProAspC 64
                  |||||||||:::::         :::::::|||
     212 GGCGGCGGTGGACGAGCTGGGCAAAGTG.........CTGACGCCCACCC 252

      64 ysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLys 80
           :::   |||           ||||||
     253 AGGTTAAGAATAGACCCACCAGCATTTCGTGG..................... 284

      81 PheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspPr 97
            |||      :::|||      |||   ||||||:::::|||||
     285 ...GATGGTCTTGATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCC 331

      97 oAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpL 114
         ||||||||||||||||   :::|||:::   |||   |||:::|||:::|
     332 GGATGCTCCCAGCAGGAAGGATCCCAAATACAGAGAATGGCATCATTTCC 381

     114 euValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGly 130
         |||||:::::::::|||||   |||:::::::::|||   :::
     382 TGGTGGTCAACATGAAGGGCAATGACATCAGCAGTGGCACAGTC...... 425

     131 GlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPh 147
            ||||||   |||   :::   :::|||||      :::|||::
     426 ......CTCTCCGATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCT 469

     147 eHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer. 163
         :|||||||||   ::::::|||||   |||:::     :::
     470 CCACCGCTATGTCTGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTG 519

     164 ..........LeuLeuProLysGluAsnLysThrArgGlySerTrpLys 176
                 |||      :::::::::      ||||||::::::|||
     520 ACGAGCCCATCCTCAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAG 569

     177 MetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerTh 193
         :::      |||   ::::::::::|||      |||   |||:::||
     570 GTGGCGTCCTTCCGTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCAC 619

     193 rGlnPheMetThrGln 198
         |   ::::::::::::
     620 GTGTTACCAGGCCGAG 635
```

```
OM of: US-09-215-435-225 to: Issued_Patents_NA:*  out_format : pfs

Date: Nov 15, 2001  4:23 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4 + 1213.00 2632.22  5.8e-139  903  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-3 +  212.50  436.77  1.1e-16   1047  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-14 +  208.50  424.32  5.5e-16   1447  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-644-372-1 +  201.50  415.40  1.7e-15   822  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-1 +  164.00  332.73  7.0e-11   855  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-3 -  164.00  332.71  7.0e-11   856  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-1021 -  122.50  243.59  6.4e-06   726  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-991 +  115.00  227.69  4.9e-05   692  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2 -  113.50  124.78  26.68     4403765
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-1 -  113.50  124.76  26.75     4411529
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-47 +   87.50  167.26  0.1148    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-11 +   87.50  167.26  0.1148    700  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-47 +   87.50  167.26  0.1148    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-8 +   87.50  145.38  1.90      4794  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-45 +   87.50  145.37  1.90      4795  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-45 +   87.50  145.37  1.90      4795  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-46 +   82.50  156.30  0.4684    700  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-46 +   82.50  156.30  0.4684    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-10 +   82.00  155.22  0.5380    699  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-07-765-830A-5 +   80.00  130.28  13.18     4258  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-288-630-1 +   79.50  142.39  2.79      1333  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-785-241-2 +   79.00  131.95  10.63     3031  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-056-105-10 +   77.50  123.21  32.62     4895  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-056-105-9 +   77.50  113.50  113.34    11495  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-814-095-7 -   77.00   99.72  663.67    35060  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-697-610-1 +   76.00  128.82  15.90     2240  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-349-357-1 +   76.00  128.82  15.90     2240  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-16980-1 +   76.00  128.23  17.15     2359
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-211-417-2 +   76.00  124.46  27.81     3286  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-770-379-19 +   76.00   97.52  880.67    35100  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-757-669A-19 +   76.00   97.52  880.67    35100
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-791-115B-8 -   75.00  133.19  9.07      1257  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-791-115B-2 -   75.00  122.71  34.80     3160  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-368-590-1 -   74.00  110.22  172.63    7812  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-812-13 +   74.00  107.98  230.18    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-827-13 +   74.00  107.98  230.18    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-921-177-13 +   74.00  107.98  230.18    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-362-577C-13 +   74.00  107.98  230.18    9515  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-920-828-13 +   74.00  107.98  230.18    9515  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-492-027A-9 +   73.50  125.00  25.94     1935  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-999-774A-5 -   73.50  120.27  47.57     2932
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-189-760-5 +   73.00  126.58  21.18     1529
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-188-811-5 +   73.00  126.58  21.18     1529
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-514-422-5 +   73.00  126.58  21.18     1529
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-656-984A-33 +   73.00  125.46  24.44     1687

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4

seq_documentation_block:
; Sequence 4, Application US/09208718
; Patent No. 6063767
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer
;    APPLICANT:  Corley, Neil
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN PHOSPHOLIPID BINDING PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,718
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/958,820
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0379 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  LUNGTUT12
;      CLONE:  3126479
US-09-208-718-4


alignment_scores:
            Quality: 1213.00              Length:    227
              Ratio:  5.367                 Gaps:      0
 Percent Similarity:  99.559   Percent Identity: 99.119

alignment_block:
 US-09-215-435-225 x US-09-208-718-4   ..

 Align seg 1/1  to: US-09-208-718-4  from: 1  to: 903

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 137

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 187
```

```
     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    188 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 237

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    238 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 287

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    288 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 337

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    338 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 387

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    388 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 437

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||:::|||||||||||||||||||||||||
    438 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 487

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    488 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 537

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    538 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 587

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    588 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCT 637

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    638 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 687

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        |||||||||||||||||||||||||||||   ||||||||||||||||||
    688 CAGGACTCACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAA 737

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    738 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 768

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-3

seq_documentation_block:
; Sequence 3, Application US/08403378B
; Patent No. 5759991
;  GENERAL INFORMATION:
;    APPLICANT:  TOHDOH, NAOKI
;    APPLICANT:  TOJO, SHIN-ICHIRO
;    APPLICANT:  KOJIMA, SHIN-ICHI
;    APPLICANT:  UEKI, YASUYUKI
;    APPLICANT:  NISHIHARA, TOSHIO
;    APPLICANT:  FUKUSHIMA, NOBUYUKI
;    APPLICANT:  IRIE, TSUNEMASA
;    APPLICANT:  ONO, KEIICHI
;    APPLICANT:  AGUI, HIDEO
;    APPLICANT:  OJIKA, KOSEI
;    TITLE OF INVENTION:  NEUROTROPHIC PEPTIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 PENNSYLVANIA AVENUE, NW
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,378B
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-124688
;      FILING DATE:  27-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-080398
;      FILING DATE:  30-MAR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-280590
;      FILING DATE:  27-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-333241
;      FILING DATE:  21-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-243003
;      FILING DATE:  12-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/758,043
;      FILING DATE:  12-SEP-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/873,764
;      FILING DATE:  27-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01214
;      FILING DATE:  27-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGART, WADDELL A
;      REGISTRATION NUMBER:  24,861
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1047 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus norvegicus
;      STRAIN:  Wistar
;      TISSUE TYPE:  hippocampal tissue of brain
;    IMMEDIATE SOURCE:
;      CLONE:  A010-12
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  26..586
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..25
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  26..586
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  587..1047
;    FEATURE:
;      NAME/KEY:  polyA_site
;      LOCATION:  1008..1047
;    FEATURE:
```

```
;       NAME/KEY:  polyA_signal
;       LOCATION:  987..992
US-08-403-378B-3


alignment_scores:
            Quality:  212.50              Length:    186
              Ratio:  2.063                Gaps:      8
 Percent Similarity:  55.376   Percent Identity: 32.796

alignment_block:
 US-09-215-435-225 x US-08-403-378B-3   ..

 Align seg 1/1  to: US-08-403-378B-3  from: 1  to: 1047

      52 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 68
         |||||||||:::::         :::::::|||              ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      68 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 85
         |   :::      ||||||                          |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      85 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 101
            |||      |||   ||||||::::::||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

     102 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 118
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

     118 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 135
         :||||||   |||::::::::::|||   :::             ||||||
     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC.............CTCTCCG 338

     135 laTyrGlnAlaProSerProProAlaHisSerGlyPheHis...ArgTyr 150
           |||   :::   |||   |||       :::|||:::|||   ||||||
     339 AATAC...GTGGGCTCCGGACCTAAAGACACAGGTCTGCACCGCCGCTAC 385

     151 GlnPheValTyrLeuGlnGluGlyLysValIleSer............ 163
         :::::|||||   ||||||      :::
     386 GTCTGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCAT 435

     164 .LeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgP 180
          |||      |||::::::   :::||||||:::::::|||::::::   |
     436 CCTCAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCT 485

     180 heLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMet 196
         ||   :::::::::|||||||||   |||   |||:::|||   |||:::
     486 TCCGCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAG 535

     197 ThrGlnAsnTyrGlnAspSer........................... 203
         ::::::   ::::::||||||
     536 GCAGAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGTCGGCTGG 582

     204 .....................................ProThrLeuGlnA 208
                                             ||||||:::|||:
     583 GAAGTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGT 632

     208 laProArg 210
         ::   |||
     633 CAAGTCGT 640

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-14

seq_documentation_block:
; Sequence 14, Application US/08403378B
; Patent No. 5759991
;  GENERAL INFORMATION:
;    APPLICANT:  TOHDOH, NAOKI
;    APPLICANT:  TOJO, SHIN-ICHIRO
;    APPLICANT:  KOJIMA, SHIN-ICHI
;    APPLICANT:  UEKI, YASUYUKI
;    APPLICANT:  NISHIHARA, TOSHIO
;    APPLICANT:  FUKUSHIMA, NOBUYUKI
;    APPLICANT:  IRIE, TSUNEMASA
;    APPLICANT:  ONO, KEIICHI
;    APPLICANT:  AGUI, HIDEO
;    APPLICANT:  OJIKA, KOSEI
;    TITLE OF INVENTION:  NEUROTROPHIC PEPTIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 PENNSYLVANIA AVENUE, NW
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,378B
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-124688
;      FILING DATE:  27-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-080398
;      FILING DATE:  30-MAR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-280590
;      FILING DATE:  27-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-333241
;      FILING DATE:  21-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-243003
;      FILING DATE:  12-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/758,043
;      FILING DATE:  12-SEP-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/873,764
;      FILING DATE:  27-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01214
;      FILING DATE:  27-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGART, WADDELL A
;      REGISTRATION NUMBER:  24,861
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  placental tissue
;    IMMEDIATE SOURCE:
;      CLONE:  p1-3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  120..680
```

```
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..680
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  681..1447
US-08-403-378B-14


alignment_scores:
            Quality:  208.50              Length:    206
              Ratio:  1.813                 Gaps:      7
 Percent Similarity:  55.825   Percent Identity: 31.553

alignment_block:
 US-09-215-435-225 x US-08-403-378B-14   ..

 Align seg 1/1  to: US-08-403-378B-14  from: 1  to: 1447

       3 TrpThrMetArgLeuValThrAlaAla.LeuLeuLeuGlyLeuMetMetV 19
         |||           :::   ||||||    ||||||||||||   |||
      74 TGGCCTACCGCGGCACTCCCGGCTGCACGCTCTGCTTGGCCTCGCCATGC 123

      19 alValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeu 35
           |||            |||   ::::::      :::      :::|||
     124 CGGTG............GACCTCAGCAAGTGGTCCGGGCCCTTGAGCCTG 161

      36 LeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr..... 50
           :::   |||           :::   |||:::|||   |||
     162 CAAGAAGTGGACGAGCAGCCGCAGCACCCGCTGCATGTCACCTACGCCGG 211

      51 ..........ProGluLeuGlyAsnIleGlyCysLysValValProAspC 64
                   |||||||||:::::        :::::::|||
     212 GGCGGCGGTGGACGAGCTGGGCAAAGTG.........CTGACGCCCACCC 252

      64 ysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLys 80
           :::   |||            ||||||
     253 AGGTTAAGAATAGACCCACCAGCATTTCGTGG.................. 284

      81 PheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspPr 97
            |||      :::|||      |||   ||||||:::::|||||
     285 ...GATGGTCTTGATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCC 331

      97 oAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpL 114
         ||||||||||||||||   :::|||:::   |||   |||:::|||:::|
     332 GGATGCTCCCAGCAGGAAGGATCCCAAATACAGAGAATGGCATCATTTCC 381

     114 euValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGly 130
         |||||:::::::::||||||   |||:::::::::|||   :::
     382 TGGTGGTCAACATGAAGGGCAATGACATCAGCAGTGGCACAGTC...... 425

     131 GlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPh 147
            ||||||   |||   :::   :::||||||      :::|||::
     426 ......CTCTCCGATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCT 469

     147 eHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer. 163
         :|||||||||   ::::::||||||   |||:::      :::
     470 CCACCGCTATGTCTGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTG 519

     164 ..........LeuLeuProLysGluAsnLysThrArgGlySerTrpLys 176
                   |||      :::::::::      ||||||::::::|||
     520 ACGAGCCCATCCTCAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAG 569

     177 MetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerTh 193
         :::      |||   ::::::::::::|||      |||   |||:::||
     570 GTGGCGTCCTTCCGTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCAC 619

     193 rGlnPheMetThrGln 198
         |   ::::::::::::
     620 GTGTTACCAGGCCGAG 635


seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-644-372-1

seq_documentation_block:
; Sequence 1, Application US/07644372
; Patent No. 5416009
;  GENERAL INFORMATION:
;    APPLICANT:  Lazzeri, Mario E.
;    APPLICANT:  Nutman, Thomas B.
;    APPLICANT:  Weiss, Niklaus
;    TITLE OF INVENTION:  A DNA SEGMENT ENCODING A SPECIFIC
;    TITLE OF INVENTION:  IMMUNODIAGNOSTIC ANTIGEN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1615 L. Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/644,372
;      FILING DATE:  19910123
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)861-3000
;      TELEFAX:  (202)822-0944
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  52..507
US-07-644-372-1


alignment_scores:
            Quality:  201.50              Length:    147
              Ratio:  2.144                 Gaps:      7
 Percent Similarity:  63.946   Percent Identity: 36.054

alignment_block:
 US-09-215-435-225 x US-07-644-372-1   ..

 Align seg 1/1  to: US-07-644-372-1  from: 1  to: 822

      55 AsnIleGlyCysLysValValPro...AspCysAsnAsnTyrArgGlnLy 70
         |||:::|||   ::::::::::|||   :::   :::|||         ||
     244 AATCTGGGCAATGAACTTACGCCGACGCAGGTAAAGAATCAGCCGACAAA 293

      70 sIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAspG 87
         |:::   ||||||                                |||      |
     294 AGTA...TCATGG.....................GATGCGGAACCTG 316

      87 lyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAla 103
         |||||   |||   |||||||||:::|||||||||||||||||||||||
     317 GAGCCTTATATACGCTCGTTATGACTGATCCGGACGCACCATCTCGAAAA 366

     104 GluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGl 120
         :::|||      |||   |||:::|||||||||:::   :::|||:::||
     367 AACCCCGTATTCAGAGAGTGGCACCATTGGTTGATAATTAATATTTCTGG 416

     120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr. 136
         |   ::::::::::::|||   :::            ||||||   |||
     417 ACAAAATGTTAGCAGTGGCACAGTG...........TTATCTGATTATT 454
```

```
     137 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 153
     455 GGATCA...GGTCCACGAAAAGGCACAGGACTTCATCGTTATGTATTCTT 501

     153 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 170
     502 GGTTTATAAACAACCTGGAAGTATCACGGAT......ACTCAACATGGCG 545

     170 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 186
     546 GAAATCGCCGAAATTTCAAAGTTATGGATTTTGCAAACAAACATCACTTG 595

     187 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsn 199
     596 GGAAATCCAGTTGCCGGAAACTTCTTCCAGGCTAAACAT 634

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-1

seq_documentation_block:
; Sequence 1, Application US/09060726A
; Patent No. 6225530
; GENERAL INFORMATION:
;  APPLICANT:  Weigel, Detlef
;  APPLICANT:  Salk Institute
;  TITLE OF INVENTION:  FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE:  SALKINS.026A
;  CURRENT APPLICATION NUMBER: US/09/060,726A
;  CURRENT FILING DATE:  1998-04-15
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 855
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (63)...(590)
US-09-060-726A-1


alignment_scores:
            Quality:  164.00             Length:    113
              Ratio:  2.412                Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225 x US-09-060-726A-1   ..

 Align seg 1/1  to: US-09-060-726A-1  from: 1  to: 855

      90 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 106
     255 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 304

     106 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 123
     305 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 350

     123 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 139
     351 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 395

     140 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 156
     396 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 439

     156 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 173
     440 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 470

     173 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 185
     471 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 518

     186 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
     519 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 557

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-3

seq_documentation_block:
; Sequence 3, Application US/09060726A
; Patent No. 6225530
; GENERAL INFORMATION:
;  APPLICANT:  Weigel, Detlef
;  APPLICANT:  Salk Institute
;  TITLE OF INVENTION:  FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE:  SALKINS.026A
;  CURRENT APPLICATION NUMBER: US/09/060,726A
;  CURRENT FILING DATE:  1998-04-15
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-060-726A-3


alignment_scores:
            Quality:  164.00             Length:    113
              Ratio:  2.412                Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225 x US-09-060-726A-3/rev   ..

 Align seg 1/1  to reverse of: US-09-060-726A-3  from: 1  to: 856

      90 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 106
     602 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 553

     106 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 123
     552 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 507

     123 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 139
     506 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 462

     140 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 156
     461 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 418

     156 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 173
     417 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 387

     173 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 185
     386 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 339

     186 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
     338 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 300

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-1021

seq_documentation_block:
; Sequence 1021, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1021:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  726 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1025UP
US-08-998-416-1021


alignment_scores:
            Quality:  122.50              Length:     89
              Ratio:  2.149                 Gaps:      3
 Percent Similarity:  64.045   Percent Identity: 34.831

alignment_block:
 US-09-215-435-225 x US-08-998-416-1021/rev   ..

 Align seg 1/1  to reverse of: US-08-998-416-1021  from: 1  to: 726

      87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
         |||      :::   |||   |||:::|||||||||||||||||||||||:: 
     717 GGGGACCTATTCACGCTGGGGATGACAGACCCAGATGCTCCCTCGCGGTC 668

     103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys. 119
         ::::   :::          :::   |||:::|||   |||:::||| 
     667 GGACCACAAGTGGTCGGAATACTGCCACTTTCTGGAAACGAACATAACGC 618

     120 ..GlyAlaAsp......LeuLysLysGlyLysIleGlnGlyGlnGluLeu 133
           |||:::|||      ::::::          :::::::|||
     617 TGGGCTCGGATGACGGGGTGTCGCACGTGGTGCTAAAGGGCACCCCGCAG 568

     134 SerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTy 150
             ::::::::::|||:::|||||||||   :::|||   ||||||||
     567 GTGGAGCACATGGGCCCTGCGCCGCCGGCCGGCACAGGGGCTCACCGGTA 518

     150 rGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProL 167
         |   ::::::::::::   |||   |||::::::          |||   :
     517 CGTGTGGTTGTTGTTTCGGCAGCCCGGGCGGCTA......GAACTGAGCG 474

     167 ysGluAsnLysThrArg 172
         ::|||:::   ||||||
     473 AGGAGGAGGTCACCCGC 457

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-991

seq_documentation_block:
; Sequence 991, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  991:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  692 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1609RP
US-08-998-416-991


alignment_scores:
            Quality:  115.00              Length:    140
              Ratio:  1.620                 Gaps:      8
 Percent Similarity:  50.714   Percent Identity: 30.000

alignment_block:
 US-09-215-435-225 x US-08-998-416-991   ..

 Align seg 1/1  to: US-08-998-416-991  from: 1  to: 692
```

```
      26 GluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAspThrLeuPh 42
         ||||||:::|||   :::|||
     336 GAAAACGGTCCATCAGGACAC..........................  356

      42 eCysGlnGlyLeuGluValPheTyrPro.................GluL 53
                 |||   |||   ||||||                     |
     357 .........CTTGTAGTGGAGTACCCGGGCGAGTCTACAGCGGTGACGC 396

      53 euGlyAsnIle............GlyCysLysValValProAspCysAsn 65
         ||||||||:::            :::   ::::::||||||:::
     397 TGGGGAACGTTATGCCTGTGGAGGCTACGCAGACGGTGCCCAAC...... 440

      66 AsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPhePr 82
                                 :::      |||:::      ||
     441 .........................CTGATGTTAATCACGACCGAGCC 463

      82 oGlyAlaVal...AspGlyAlaThrTyrIleLeuValMetValAspProA 98
         ||||   |||   :::|||      :::   |||:::|||:::|||||||
     464 GGGAATCGTCAGGGAGGGGGACCTATTCACGCTGGCGATGACAGACCCAG 513

      98 spAlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeu 114
         ||||||||||||||::::::   :::        :::   |||:::|||
     514 ATGCTCCCTCGCGGTCGGACCACAAGTGGTCGGAATACTGCCACTTTCTG 563

     115 ValThrAspIleLys...GlyAlaAsp......LeuLysLysGlyLysIl 128
            |||:::|||      |||:::|||      ::::::         ::
     564 GAAACGAACATAACGCTGGGCTCGGATGACGGGGTGTCGCACGTGGTGCT 613

     128 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPro.AlaHis 144
         ::::|||               :::::::::|||:::|||||| |||
     614 AAAGGGCACCCCGCAGGTGGAGCACATGGGCCCTGCGCCGCCGGGCCGGC 663

     145 SerGlyPheHisArgTyr 150
         :::|||   |||||||||
     664 ACAGGGGCTCACCGGTAC 681

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
           Quality:  113.50            Length:    179
             Ratio:  1.419               Gaps:     11
 Percent Similarity:  44.693   Percent Identity: 27.374

alignment_block:
 US-09-215-435-225 x US-09-103-840A-2/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-2  from: 1  to: 4403765

      10 AlaAlaLeuLeuLeuGlyLeuMetMetValVal..................  20
         |||      :::|||||||||   :::   |||
 2154533GCACTGGCCATCTTGGGGCTGGCGCTCCCCGTGGCGCTAGTTGCCTACGG 2154484

      21 ....ThrGlyAspGluAspGluAsnSerProCysAlaHisGluAla....  34
             :::|||||:::   :::   :::|||   |||   :::|||
 2154483TGGCAACGGTGACAGTCGAAAGGCGGCGCCGCTGGCGCCGAAAGCAGCAG 2154434

      35 .....................................LeuLeuAsp  37
                                              |||
 2154433CGCTCGGTCGGAGTATGCCCGAAACGCCTACCGGCGATGTACTGACAATC 2154384

      38 GluAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGl  54
         ::::::      |||:::::::|||      :::      ||||||
 2154383AGCAGTCCGGCATTCGCCGACGGTGCGCCGATC......CCGGAACAGTA 2154340

      54 yAsnIle...GlyCysLysValValProAspCysAsnAsnTyrArgGlnL  70
         :::      |||:::::::::::::|||
 2154339CACCTGCAAAGGAGCCAATATCGCGCCTCCG..................... 2154309

      70 ysIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAsp  86
                 :::|||      |||          |||   ||||||
 2154308.....TTGACCTGGTCGGCGCCG.........TTTGGCGGCGCA...... 2154279

      87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
                        ||||||:::   ||||||||||||||||
 2154278..............CTCGTTGTCGATGATCCGGACGCACCT......CG 2154250

     103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys. 119
           ||||||             :::   ||||||:::|||      |||
 2154249CGAACCT............TACGTCCATTGGATCGTGATCGGGATCGCCC 2154212

     120 .....GlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSer 134
              :::::::::          |||:::      |||      :::|||
 2154211CTGGTGCTGGCAGCACCGCCGATGGTGAGACTCCCGGTGGCGGAATCAGC 2154162

     135 .......................AlaTyrGlnAlaProSerProProAl 143
                                 ||||||   :::|||   ||||||||
 2154161CTGCCGAACTCCAGCGGTCAGCCCGCATACACCGGCCCCTGCCCGCCGGC 2154112

     143 aHisSerGlyPheHisArgTyrGlnPhePheValTyr 155
         |   :::|||   |||:::|||:::|||   :::|||
 2154111GGGCACCGGGACACACCACTACCGGTTTACCCTCTAC 2154075

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


alignment_scores:
           Quality:  113.50            Length:    179
             Ratio:  1.419               Gaps:     11
```

```
 Percent Similarity:  44.693   Percent Identity: 27.374

alignment_block:
 US-09-215-435-225 x US-09-103-840A-1/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-1  from: 1  to: 4411529

     10 AlaAlaLeuLeuLeuGlyLeuMetMetValVal..... 20
        |||      :::||||||||||   :::   |||
 2157273GCACTGGCCATCTTGGGGCTGGCGCTCCCCGTGGCGCTAGTTGCCTACGG 2157224

     21 ....ThrGlyAspGluAspGluAsnSerProCysAlaHisGluAla.... 34
           :::|||||||:::   :::   :::|||   |||   :::|||
 2157223TGGCAACGGTGACAGTCGAAAGGCGGCGCCGCTGGCGCCGAAAGCAGCAG 2157174

     35 .....LeuLeuAsp 37
                  |||
 2157173CGCTCGGTCGGAGTATGCCCGAAACGCCTACCGGCGATGTACTGACAATC 2157124

     38 GluAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGl 54
        ::::::      |||:::::|||      :::      ||||||
 2157123AGCAGTCCGGCATTCGCCGACGGTGCGCCGATC......CCGGAACAGTA 2157080

     54 yAsnIle...GlyCysLysValValProAspCysAsnAsnTyrArgGlnL 70
        :::      |||:::::::::::::|||
 2157079CACCTGCAAAGGAGCCAATATCGCGCCTCCG..... 2157049

     70 ysIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAsp 86
              :::|||      |||         |||   ||||||
 2157048.....TTGACCTGGTCGGCGCCG.....TTTGGCGGCGCA..... 2157019

     87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
                       |||||||:::   |||||||||||||||
 2157018.............CTCGTTGTCGATGATCCGGACGCACCT......CG 2156990

    103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys. 119
        ||||||            :::   ||||||:::|||      |||
 2156989CGAACCT............TACGTCCATTGGATCGTGATCGGGATCGCCC 2156952

    120 .....GlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSer 134
             :::::::::        |||:::      |||      :::|||
 2156951CTGGTGCTGGCAGCACCGCCGATGGTGAGACTCCCGGTGGCGGAATCAGC 2156902

    135 .....AlaTyrGlnAlaProSerProProAl 143
                      ||||||   :::|||   ||||||||
 2156901CTGCCGAACTCCAGCGGTCAGCCCGCATACACCGGCCCCTGCCCGCCGGC 2156852

    143 aHisSerGlyPheHisArgTyrGlnPhePheValTyr 155
        |   :::|||   |||:::|||:::|||   :::|||
 2156851GGGCACCGGGACACACCACTACCGGTTTACCCTCTAC 2156815

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-47

seq_documentation_block:
; Sequence 47, Application US/08313185
; Patent No. 5851763
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE.FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,185
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-313-185-47


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225 x US-08-313-185-47   ..

 Align seg 1/1  to: US-08-313-185-47  from: 1  to: 700

     64 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 79
        |||||||:::      :::      |||   |||      |||:::::::
    321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

     80 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 96
         :::   |||:::|||   :::   |||         |||||||:::   |
    368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.....CTCGTCGTCGATG 408

     96 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 112
        ||||||||||  :::                          ::::::   |||
    409 ACCCCGACGCGGTCGGC.....GGACTGTACGTGCAC 440

    113 TrpLeuValThrAspIle.....LysGlyAlaAsp..... 122
        |||:::|||||   |||         :::   ||||||
    441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

    123 .....LeuLysLysGlyLysIleGlnGlyG 131
                              :::::::|||      ||||||
    491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

    131 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 147
                     |||   :::|||   |||||||||   :::|||
    540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

    148 HisArgTyrGlnPhePheValTyr 155
        |||:::|||:::|||   :::|||
    576 CACCACTACCGGTTTACCCTCTAC 599

 seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-11

 seq_documentation_block:
 ; Sequence 11, Application US/08459499
 ; Patent No. 5871912
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Heym, Beate
```

```
;    APPLICANT:  Cole, Stewart T.
;    APPLICANT:  Young, Douglas B.
;    APPLICANT:  Zhang, Ying
;    TITLE OF INVENTION:  Nucleic Acid Probes, Sequences, and Methods
;    TITLE OF INVENTION:  for Detecting Mycobacterium Tuberculosis Resistant to Isoniazi
;    TITLE OF INVENTION:  Amended)
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,499
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/875,940
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/929,206
;      FILING DATE:  27-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,655
;      FILING DATE:  11-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0110-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-459-499-11


alignment_scores:
            Quality:   87.50            Length:    108
              Ratio:  1.620              Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225 x US-08-459-499-11   ..

 Align seg 1/1  to: US-08-459-499-11  from: 1  to: 700

      64 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 79
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      80 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 96
          :::   |||:::|||   :::  |||           |||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      96 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 112
         ||||||||||   :::                    ::::::   |||
     409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

     113 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 122

         |||:::||||||   |||               :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     123 .......................LeuLysLysGlyLysIleGlnGlyG 131
                                      :::::|||      ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     131 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 147
                      |||   :::|||   |||||||||   :::|.||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     148 HisArgTyrGlnPhePheValTyr 155
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-47

seq_documentation_block:
; Sequence 47, Application US/09082614A
; Patent No. 6124098
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/082,614A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/313,185
;      FILING DATE:  12-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-082-614A-47


alignment_scores:
            Quality:   87.50            Length:    108
              Ratio:  1.620              Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481
```

```
alignment_block:
 US-09-215-435-225 x US-09-082-614A-47   ..

 Align seg 1/1  to: US-09-082-614A-47  from: 1  to: 700

      64 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 79
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      80 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 96
          :::   |||:::|||   :::   |||         ||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      96 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 112
         ||||||||||   :::                    ::::::   |||
     409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

     113 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 122
         |||:::||||||   |||                :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     123 .......................LeuLysLysGlyLysIleGlnGlyG 131
                                :::::::|||      ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     131 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 147
                   |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     148 HisArgTyrGlnPhePheValTyr 155
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-8

seq_documentation_block:
; Sequence 8, Application US/08459499
; Patent No. 5871912
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart T.
;    APPLICANT:  Young, Douglas B.
;    APPLICANT:  Zhang, Ying
;    TITLE OF INVENTION:  Nucleic Acid Probes, Sequences, and Methods
;    TITLE OF INVENTION:  for Detecting Mycobacterium Tuberculosis Resistant to Isoniazi
;    TITLE OF INVENTION:  Amended)
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,499
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/875,940
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/929,206
;      FILING DATE:  27-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,655
;      FILING DATE:  11-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0110-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4794 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-459-499-8


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225 x US-08-459-499-8   ..

 Align seg 1/1  to: US-08-459-499-8  from: 1  to: 4794

      64 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 79
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      80 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 96
          :::   |||:::|||   :::   |||         ||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      96 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 112
         ||||||||||   :::                    ::::::   |||
     409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

     113 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 122
         |||:::||||||   |||                :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     123 .......................LeuLysLysGlyLysIleGlnGlyG 131
                                :::::::|||      ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     131 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 147
                   |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     148 HisArgTyrGlnPhePheValTyr 155
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-45

seq_documentation_block:
; Sequence 45, Application US/08313185
; Patent No. 5851763
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
```

```
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,185
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4795 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-313-185-45


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225 x US-08-313-185-45   ..

 Align seg 1/1  to: US-08-313-185-45  from: 1  to: 4795

      64 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 79
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      80 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 96
          :::   |||:::|||   :::   |||         ||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      96 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 112
         ||||||||||   :::                  ::::::   |||
     409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

     113 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 122
         |||:::||||||   |||               :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     123 .......................LeuLysLysGlyLysIleGlnGlyG 131
                               ::::::|||      ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     131 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 147
                   |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     148 HisArgTyrGlnPheValTyr 155
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-124 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-471-276-7 + 1224.00 2268.67  5.1e-117  826  !
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-7 + 1224.00 2268.67  5.1e-117  826  !
/cgn1_7/ptodata/1/pna/US6008_COMB.seq:US-60-081-563-70 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-7 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-7 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-7 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-7 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-224-006-7 + 1224.00 2268.67  5.1e-117  826
/cgn1_7/ptodata/1/pna/US6027_COMB.seq:US-60-278-258-12888 + 1224.00 2267.80  5.7e-117  9
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-726-176-1502 + 1220.00 2260.70  1.4e-116  87
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-56 + 1218.00 2255.37  2.8e-116  1028
/cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-56 + 1218.00 2255.37  2.8e-116  1028
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-56 + 1218.00 2255.37  2.8e-116  1028
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-10894 + 1217.00 2254.74  3.0e-116  9
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-922-5157 + 1213.00 2247.55  7.6e-116  878
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-922-5157 + 1213.00 2247.55  7.6e-116  878
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-958-820-4 + 1213.00 2247.28  7.9e-116  903  !
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-390-126-4 + 1213.00 2247.28  7.9e-116  903  !
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-61 + 1198.00 2218.78  3.0e-114  952
/cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-61 + 1198.00 2218.78  3.0e-114  952
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-61 + 1198.00 2218.78  3.0e-114  952
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-931-836-60 + 1189.00 2202.63  2.4e-113  890
/cgn1_7/ptodata/1/pna/US096A_COMB.seq:US-09-617-746A-9 + 1185.00 2194.31  7.0e-113  972
/cgn1_7/ptodata/1/pna/US096B_COMB.seq:US-09-631-451A-9 + 1185.00 2194.31  7.0e-113  972
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-7851 + 1167.00 2161.42  4.8e-111  90
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-491-404-1525 + 1155.00 2139.21  8.2e-110  886
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-922-279-1525 + 1155.00 2139.21  8.2e-110  886
/cgn1_7/ptodata/1/pna/US6028_COMB.seq:US-60-281-593-409 +  844.00 1562.25  1.1e-77   615
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-2551 +  788.00 1458.84  6.5e-72   54
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-2655 +  676.00 1251.15  2.4e-60   476
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-144-182 +  652.00 1206.96  6.9e-58   447
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-3691 +  644.00 1193.24  4.0e-57   394
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-2207 +  612.00 1133.18  8.9e-54   408
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163A-4626 +  604.50 1117.54  6.6e-53   48
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163B-4626 +  604.50 1117.54  6.6e-53   48
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163D-4626 +  604.50 1117.54  6.6e-53   48
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-840-145-4626 +  604.50 1117.54  6.6e-53   484
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4958 +  578.00 1069.37  3.2e-50   423
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4958 +  578.00 1069.37  3.2e-50   42
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-4300 +  570.00 1056.10  1.8e-49
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-4523 +  565.00 1046.85  5.8e-49
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-824-559-6494 +  558.00 1031.93  3.9e-48
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-235-076-5859 +  542.50 1002.98  1.6e-46
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-248-797-14649 +  542.50 1002.98  1.6e-46

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-124

seq_documentation_block:
; Sequence 124, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 124
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 795..800
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 814..826
US-09-215-435-124


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-09-215-435-124   ..

 Align seg 1/1  to: US-09-215-435-124  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

  67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

  84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

 101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

 117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

 134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

 151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

 167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

 184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

 201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

 217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
     ||||||||||||||||||||||||||||||
 665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695
```

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-471-276-7

seq_documentation_block:
; Sequence 7, Application US/09471276
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: GENSET.025CP1
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..695
; NAME/KEY: sig_peptide
; LOCATION: 15..80
; OTHER INFORMATION: score 8.5
; OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-09-471-276-7

alignment_scores:
Quality: 1224.00 Length: 227
Ratio: 5.392 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-225 x US-09-471-276-7 ..

Align seg 1/1 to: US-09-471-276-7 from: 1 to: 826

```
   1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

  17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

  34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

  51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

  67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

  84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

 101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

 117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

 134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

 151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

 167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

 184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

 201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

 217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
     ||||||||||||||||||||||||||||||
 665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695
```

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-7

seq_documentation_block:

```
; Sequence 7, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-09-834-366-7


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-09-834-366-7    ..

 Align seg 1/1  to: US-09-834-366-7  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:US-60-081-563-70

seq_documentation_block:
; Sequence 70, Application US/60081563
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/081,563
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.027PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  70:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  826 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Prostate
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..695
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  15..80
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 8.5
```

```
;       OTHER INFORMATION:  seq AALLLGLMMVVTG/DE
;     FEATURE:
;       NAME/KEY:  polya_signal
;       LOCATION:  795..800
;     FEATURE:
;       NAME/KEY:  poly_a
;       LOCATION:  814..826
;       IDENTIFICATION METHOD:  blastn
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  426..664
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 99
;       OTHER INFORMATION:  region 99..337
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  327..435
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 1..109
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  716..777
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 100
;       OTHER INFORMATION:  region 391..452
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  660..720
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 334..394
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  506..630
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 90
;       OTHER INFORMATION:  region 327..451
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  323..435
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 94
;       OTHER INFORMATION:  region 143..255
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  180..280
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 2..102
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  464..528
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 284..348
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  279..321
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 100
;       OTHER INFORMATION:  region 100..142
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  426..466
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 97
;       OTHER INFORMATION:  region 245..285
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  358..674
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 141..457
;       OTHER INFORMATION:  id AA192427
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  670..779
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 37..146
;       OTHER INFORMATION:  id AA192427
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  485..602
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 224..341
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  601..709
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 91
;       OTHER INFORMATION:  region 115..223
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  704..779
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 46..121
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  392..436
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 97
;       OTHER INFORMATION:  region 393..437
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  429..465
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 94
;       OTHER INFORMATION:  region 363..399
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  464..493
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 334..363
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  343..730
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 8..395
;       OTHER INFORMATION:  id HSPD04880
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  507..663
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
```

```
;      OTHER INFORMATION:  region 264..420
;      OTHER INFORMATION:  id AA192426
US-60-081-563-70


alignment_scores:
           Quality: 1224.00              Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-081-563-70   ..

 Align seg 1/1  to: US-60-081-563-70  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-7

seq_documentation_block:
; Sequence 7, Application US/60147499
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Jobert, S.
;  APPLICANT: Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/60/147,499
;  CURRENT FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-147-499-7


alignment_scores:
           Quality: 1224.00              Length:    227
             Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-147-499-7   ..

 Align seg 1/1  to: US-60-147-499-7  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464
```

```
    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-7

seq_documentation_block:
; Sequence 7, Application US/60169629
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: GENSET.071PRF
;  CURRENT APPLICATION NUMBER: US/60/169,629
;  CURRENT FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-169-629-7


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-169-629-7    ..

 Align seg 1/1  to: US-60-169-629-7  from: 1  to: 826

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-7

seq_documentation_block:
; Sequence 7, Application US/60187470
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: 78.US2.PRO
;  CURRENT APPLICATION NUMBER: US/60/187,470
;  CURRENT FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-187-470-7
```

```
alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-187-470-7   ..

 Align seg 1/1  to: US-60-187-470-7  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-7

seq_documentation_block:
; Sequence 7, Application US/60197873
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
;  FILE REFERENCE:  81.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/197,873
;  CURRENT FILING DATE:  2000-04-18
;  NUMBER OF SEQ ID NOS:  52153
;  SOFTWARE:  Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-197-873-7


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-197-873-7   ..

 Align seg 1/1  to: US-60-197-873-7  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-224-006-7

seq_documentation_block:
; Sequence 7, Application US/60224006
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  TITLE OF INVENTION:  ESTs encoding human proteins
;  FILE REFERENCE:  95.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/224,006
;  CURRENT FILING DATE:  2000-08-04
;  NUMBER OF SEQ ID NOS:  67
;  SOFTWARE:  Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-224-006-7


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-224-006-7   ..

 Align seg 1/1  to: US-60-224-006-7  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 114

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CCCTCCTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 164

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     165 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 214

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 264

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 314

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 364

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     365 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 414

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     415 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 464

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CAGTTCTTTGTCTATCTTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAA 514

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     515 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 564

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 614

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     615 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 664

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     665 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:US-60-278-258-12888

seq_documentation_block:
; Sequence 12888, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms Identified Thereby
;  FILE REFERENCE: GX-0010-1 P
;  CURRENT APPLICATION NUMBER: US/60/278,258
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 17730
;  SOFTWARE: PERL Program
; SEQ ID NO 12888
;   LENGTH: 904
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 204392.4
US-60-278-258-12888


alignment_scores:
            Quality: 1224.00              Length:    227
              Ratio:  5.392                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x US-60-278-258-12888   ..
```

```
Align seg 1/1  to: US-60-278-258-12888  from: 1  to: 904

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     89 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 138

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    139 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 188

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    189 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 238

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    239 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 288

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    289 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 338

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    339 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 388

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    389 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 438

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    439 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 488

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    489 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 538

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    539 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 588

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    589 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 638

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    639 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 688

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    689 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 738

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
    739 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 769

seq_name: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-726-176-1502

seq_documentation_block:
; Sequence 1502, Application US/09726176
; GENERAL INFORMATION:
;  APPLICANT:  Gearing, David P.
;  APPLICANT:  Fraser, Christopher C.
;  APPLICANT:  Donovan, Michael J.
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.2018-001
;  CURRENT APPLICATION NUMBER: US/09/726,176
;  CURRENT FILING DATE:  2000-11-29
;  PRIOR APPLICATION NUMBER:  60/167,841
;  PRIOR FILING DATE:  1999-11-29
;  NUMBER OF SEQ ID NOS:  1882
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1502
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-726-176-1502


alignment_scores:
            Quality: 1220.00              Length:    227
              Ratio:  5.374                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x US-09-726-176-1502   ..

 Align seg 1/1  to: US-09-726-176-1502  from: 1  to: 870

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     71 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 120

     17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    121 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 170

     34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    171 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 220

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    221 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 270

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    271 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 320

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    321 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 370

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    371 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 420

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||:::||||||||||||||||||||||||||
    421 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 470

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    471 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 520

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    521 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 570

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    571 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 620

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    621 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 670

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    671 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 720

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||||||||||
```

```
     721 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 751

seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-56

seq_documentation_block:
; Sequence 56, Application PC/TUS9929950
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/29950
;  CURRENT FILING DATE: 1999-12-16
;  EARLIER APPLICATION NUMBER: 60/112,809
;  EARLIER FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/113,006
;  EARLIER FILING DATE: 1998-12-18
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1026)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1027)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1028)
;   OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-29950-56


alignment_scores:
            Quality: 1218.00            Length:    227
              Ratio:  5.389              Gaps:        0
 Percent Similarity:  99.559   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x PCT-US99-29950-56   ..

 Align seg 1/1  to: PCT-US99-29950-56  from: 1  to: 1028

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 202

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 252

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 302

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 352

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 402

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     403 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 452

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 502

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||    ||||||||||||||||||||||||
     503 TATCAAGGGCGCCGACCTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTAT 552

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     553 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 602

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     603 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 652

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     653 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 702

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     703 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 752

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     753 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 802

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     803 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 833

seq_name: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-56

seq_documentation_block:
; Sequence 56, Application US/09591316
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035P1
;  CURRENT APPLICATION NUMBER: US/09/591,316
;  CURRENT FILING DATE: 2000-06-09
;  EARLIER APPLICATION NUMBER: PCT/US99/29950
;  EARLIER FILING DATE: 1999-12-16
;  EARLIER APPLICATION NUMBER: 60/113,006
;  EARLIER FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 60/112,809
;  EARLIER FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1026)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1027)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1028)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-591-316-56
```

```
alignment_scores:
            Quality: 1218.00              Length:    227
              Ratio:  5.389                Gaps:      0
 Percent Similarity:  99.559    Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x US-09-591-316-56   ..

 Align seg 1/1  to: US-09-591-316-56  from: 1  to: 1028

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 202

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 252

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 302

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 352

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 402

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 452

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 502

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||   ||||||||||||||||||||||||||
     503 TATCAAGGGCGCCGACCTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTAT 552

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     553 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 602

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     603 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 652

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     653 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 702

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     703 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 752

     201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     753 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 802

     217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||||||||||
     803 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 833

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-56

seq_documentation_block:
; Sequence 56, Application US/09895298
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035P1
;  CURRENT APPLICATION NUMBER: US/09/895,298
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: 09/591,16
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: PCT/US99/29950
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: 60/113,006
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 60/112,809
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1026)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1027)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1028)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-895-298-56


alignment_scores:
            Quality: 1218.00              Length:    227
              Ratio:  5.389                Gaps:      0
 Percent Similarity:  99.559    Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x US-09-895-298-56   ..

 Align seg 1/1  to: US-09-895-298-56  from: 1  to: 1028

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 202

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 252

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 302

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 352

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 402

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 452

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 502
```

```
    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||   ||||||||||||||||||||||||||
    503 TATCAAGGGCGCCGACCTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTAT 552

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    553 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 602

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    603 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 652

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    653 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTT 702

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    703 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 752

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    753 CAGGACTCACCAACCCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAA 802

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        ||||||||||||||||||||||||||||||
    803 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 833

seq_name: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-10894

seq_documentation_block:
; Sequence 10894, Application US/60172373
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT: Lal, Preeti
;  APPLICANT: Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorp
;  FILE REFERENCE: GX-0006 P
;  CURRENT APPLICATION NUMBER: US/60/172,373
;  CURRENT FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 25,772
;  SOFTWARE: PERL Program
; SEQ ID NO 10894
;   LENGTH: 904
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 204392.1
US-60-172-373-10894


alignment_scores:
            Quality: 1217.00              Length:    227
              Ratio:  5.385                Gaps:      0
 Percent Similarity:  99.559   Percent Identity: 99.559

alignment_block:
 US-09-215-435-225 x US-60-172-373-10894    ..

 Align seg 1/1  to: US-60-172-373-10894  from: 1  to: 904

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      89 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 138

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     139 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 188

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||

    189 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 238

     51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    239 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 288

     67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    289 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 338

     84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    339 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 388

    101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    389 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 438

    117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    439 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 488

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    489 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 538

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    539 CAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAA 588

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    589 GGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCT 638

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    639 TCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTAC 688

    201 GlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluProLy 217
        ||||||||||||||||||||||||||||||   ||||||||||||||||||
    689 CAGGACTCACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAA 738

    217 sHisLysAsnGlnAlaGluIleAlaAlaCys 227
        ||||||||||||||||||||||||||||||
    739 GCACAAAAACCAGGCGGAGATAGCTGCCTGC 769
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-253-167 +  208.50  444.95  1.4e-17   1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-1 +  164.00  348.66  3.3e-12   856
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-3 -  164.00  348.66  3.3e-12   856
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4131 +  158.00  334.35  2.1e-11
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-971 +  105.50  219.46  5.2e-05   5
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1413 +  103.00  219.83  5.0e-05
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3078 +   96.50  196.81  0.0010
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4132 +   92.50  185.96  0.0038
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6698 +   80.50  154.93  0.2048
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-6710 +   79.00  160.75  0.0970
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-481 -   79.00  112.24  48.92     2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3027 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3028 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3029 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3030 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3031 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3032 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3033 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3034 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3035 -   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3036 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3037 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3038 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3039 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3040 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3041 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3042 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3043 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3044 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3045 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3046 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3047 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3048 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3049 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3050 +   79.00  112.24  48.92
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3678 +   78.00  153.18  0.2565
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-330 -   77.00  115.73  31.25     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2244 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2245 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2246 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2247 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2248 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2249 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2250 -   77.00  115.73  31.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2251 -   77.00  115.73  31.25

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-253-167

seq_documentation_block:
; Sequence 167, Application US/09960253
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Lodes, Michael J.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.556
;  CURRENT APPLICATION NUMBER: US/09/960,253
;  CURRENT FILING DATE:  2001-09-20
;  NUMBER OF SEQ ID NOS: 187
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 1444
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-960-253-167


alignment_scores:
           Quality:  208.50              Length:    206
             Ratio:  1.813                 Gaps:      7
 Percent Similarity:  55.825   Percent Identity: 31.553

alignment_block:
 US-09-215-435-225 x US-09-960-253-167   ..

 Align seg 1/1  to: US-09-960-253-167  from: 1  to: 1444

       3 TrpThrMetArgLeuValThrAlaAla.LeuLeuLeuGlyLeuMetMetV 19
         |||           :::   ||||||    ||||||||||||   |||
      65 TGGCCTACCGCGGCACTCCCGGCTGCACGCTCTGCTTGGCCTCGCCATGC 114

      19 alValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeu 35
           |||           |||   ::::::      :::      :::|||
     115 CGGTG...........GACCTCAGCAAGTGGTCCGGGCCCTTGAGCCTG.152

      36 LeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr..... 50
           :::   |||           :::   |||:::|||   |||
     153 CAAGAAGTGGACGAGCAGCCGCAGCACCCGCTGCATGTCACCTACGCCGG 202

      51 ..........ProGluLeuGlyAsnIleGlyCysLysValValProAspC 64
                  |||||||||::::::         ::::::|||
     203 GGCGGCGGTGGACGAGCTGGGCAAAGTG.........CTGACGCCCACCC 243

      64 ysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLys 80
           :::   |||           ||||||
     244 AGGTTAAGAATAGACCCACCAGCATTTCGTGG.................. 275

      81 PheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspPr 97
           |||      :::|||      |||   ||||||::::::|||||
     276 ...GATGGTCTTGATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCC 322

      97 oAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpL 114
         ||||||||||||||||   :::|||:::   |||   |||:::|||:::|
     323 GGATGCTCCCAGCAGGAAGGATCCCAAATACAGAGAATGGCATCATTTCC 372

     114 euValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGly 130
         |||||:::::::::||||||   |||:::::::::|||   :::
     373 TGGTGGTCAACATGAAGGGCAATGACATCAGCAGTGGCACAGTC...... 416

     131 GlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPh 147
           ||||||   |||   :::   :::||||||    .  :::|||::
     417 ......CTCTCCGATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCT 460
```

```
     147 eHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer. 163
         :|||||||||   :::::|||||   |||:::     :::
     461 CCACCGCTATGTCTGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTG 510

     164 ..........LeuLeuProLysGluAsnLysThrArgGlySerTrpLys 176
                   |||      :::::::::      ||||||::::::|||
     511 ACGAGCCCATCCTCAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAG 560

     177 MetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerTh 193
         :::      |||   :::::::::::|||      |||   |||::::||
     561 GTGGCGTCCTTCCGTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCAC 610

     193 rGlnPheMetThrGln 198
         |   ::::::::::::
     611 GTGTTACCAGGCCGAG 626

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-1

seq_documentation_block:
; Sequence 1, Application US/09845849
; GENERAL INFORMATION:
;  APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  WEIGEL, Detlef
;  APPLICANT:  KARDAILSKY, Igor
;  TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE: SALKINS.026DV1
;  CURRENT APPLICATION NUMBER: US/09/845,849
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/060,726
;  PRIOR FILING DATE: 1998-04-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-845-849-1


alignment_scores:
            Quality:  164.00            Length:    113
              Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225 x US-09-845-849-1   ..

 Align seg 1/1  to: US-09-845-849-1  from: 1  to: 856

      90 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 106
         |||   ||||||||||||||||||||||:::|||||||   :::::::::: 
     255 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 304

     106 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 123
         :   |||   :::   ||||||||||||||||||||   ::::::
     305 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 350

     123 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 139
              |||         |||:::|||:::   :::|||:::   |||
     351 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 395

     140 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 156
         ||||||   |||      |||:::||||||      |||:::::::::
     396 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 439

     156 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 173
          |||   |||:::              |||
     440 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 470

     173 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 185
           :::|||:::::::::            :::|||   :::   ::::::

     471 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 518

     186 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
         ||||||   |||   |||::::::   :::      |||
     519 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 557

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-3

seq_documentation_block:
; Sequence 3, Application US/09845849
; GENERAL INFORMATION:
;  APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  WEIGEL, Detlef
;  APPLICANT:  KARDAILSKY, Igor
;  TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE: SALKINS.026DV1
;  CURRENT APPLICATION NUMBER: US/09/845,849
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/060,726
;  PRIOR FILING DATE: 1998-04-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-845-849-3


alignment_scores:
            Quality:  164.00            Length:    113
              Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225 x US-09-845-849-3/rev   ..

 Align seg 1/1  to reverse of: US-09-845-849-3  from: 1  to: 856

      90 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 106
         |||   ||||||||||||||||||||||:::|||||||   :::::::::: 
     602 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 553

     106 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 123
         :   |||   :::   ||||||||||||||||||||   :::::::
     552 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 507

     123 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 139
              |||         |||:::|||:::   :::|||:::   |||
     506 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 462

     140 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 156
         ||||||   |||      |||:::||||||      |||:::::::::
     461 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 418

     156 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 173
          |||   |||:::              |||
     417 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 387

     173 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 185
           :::|||:::::::::            :::|||   :::   ::::::
     386 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 339

     186 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
         ||||||   |||   |||::::::   :::      |||
     338 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 300

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4131

seq_documentation_block:
; Sequence 4131, Application US/60325448
; GENERAL INFORMATION:
```

```
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 4131
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-4131


alignment_scores:
           Quality:  158.00            Length:    141
             Ratio:  1.859               Gaps:      4
 Percent Similarity:  60.284   Percent Identity: 29.078

alignment_block:
 US-09-215-435-225 x US-60-325-448-4131   ..

 Align seg 1/1  to: US-60-325-448-4131  from: 1  to: 891

      76 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 92
      255 AAGCCACGCATTGAAGTCCAGGGTGGGGACATGAGATCCTTCTTCACTCT 304

      92 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 109
      305 GGTCATGACGGACCCAGATGTGCCAGGGCCTAGTGATCCATACCTTAGGG 354

      109 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 125
      355 AGCATCTCCACTGGATCGTCAGTGATATTCCTGGCACC.........ACA 395

      126 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 142
      396 GATGCTTCTTTTGGAAGGGAGGTGGTGAGCTACGAGAGCCCCAAGCCC.. 443

      142 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 159
      444 ....AACATTGGCATTCACAGGTTCACCTTCGTGCTGTTCCAACAGAAGA 489

      159 lyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrp 175
      490 AGCGGCAGGCCATGAACCCTCCC.........TCCACCAGGGACTACTTC 530

      176 LysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSe 192
      531 AACACGCGCCGCTTCGCCAACGAGAACGACCTTGGCCTCCCGGTGGCTGC 580

      192 rThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaP 209
      581 CGTCTACTTCAACGCGCAG......CGGGAGACAGCGGCACGCCGCCGCT 624

      209 roArgGluArgAlaSerGluPro 216
      625 GACGGAAACAATCAGCCCAACCC 647

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-971

seq_documentation_block:
; Sequence 971, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 971
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-971


alignment_scores:
           Quality:  105.50            Length:    113
             Ratio:  1.598               Gaps:      2
 Percent Similarity:  58.407   Percent Identity: 26.549

alignment_block:
 US-09-215-435-225 x US-60-325-448-971   ..

 Align seg 1/1  to: US-60-325-448-971  from: 1  to: 535

      86 AspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerAr 102
      185 GATCTTTCTTCACATTGGTGGGTTATGACAGACCCAGATGTGCCAGGACC 234

      102 gAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleL 119
      235 AAGTGATCCTTATCTAAGGGAGCACCTTCATTGCATTTACGCAGGATTGT 284

      119 ysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAla 135
      285 TACTGATATACCTGGGACAACGGATGCTTCTTTTGCGCGAGGTCATAAGC 334

      136 TyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPh 152
      335 TATGAGAGTCCAAAGCCG......AACATTGGCATCCATAGGTTCATTTT 378

      152 ePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluA 169
      379 TGTGCTCTTCAAGCAGAAGCGCAGGCAAACTGTAATTGTGCCA....... 421

      169 snLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHis 185
      422 ..TCCTTCAGGGACCATTTCAACACCCGCCGGTTCGCCGAGGAGAATGAT 469

      186 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
      470 CTTGGCCTTCCTGTGGCTGCTGTCTACTTCAATGCCCAG 508

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1413

seq_documentation_block:
; Sequence 1413, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 1413
;   LENGTH: 321
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-1413


alignment_scores:
           Quality:  103.00            Length:    107
             Ratio:  1.537               Gaps:      4
 Percent Similarity:  62.617   Percent Identity: 30.841

alignment_block:
 US-09-215-435-225 x US-60-325-448-1413   ..

 Align seg 1/1  to: US-60-325-448-1413  from: 1  to: 321
```

```
     93 ValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgPh 109
        ||||||:::|||||||||:::|||:::   :::::::|||      |||
      1 GTTATGACAGACCCAGATGTGCCAGGACCAAGTGATCCATATCTAAGGGA 50

    109 eTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp.LeuLysLys 125
            |||   :::||||||||||||   |||:::    :::
     51 ACACCTACATTG.ATTGTAACTGATATACCTGGAACAACGGATG...... 93

    126 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 142
          :::   |||:::|||:::   :::|||:::::|||:::|||
     94 ...CCTCTTTTGGGACGGGAAATCATAAGCTATGAGAGCCCAAAGCCCAG 140

    142 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 159
              |||:::||||||:::   |||   ::::::   |||:::
    141 CATT......GGTATCCACAGGTTCGTTTTTGTGCTCTTCAAGCAGAAGC 184

    159 lyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrp 175
        :::       ::::::|||          :::::::|||      :::
    185 GTAGGCAGGCTGTAGTTGTGCCA.........TCCTCTAGGGATCATTTC 225

    176 LysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSe 192
        :::      :::|||   ::::::   :::||||||   |||   |||::
    226 AATACACGCCAGTTTGCTGAGGAGAACGAACTTGGCCTTCCTGTCGCTGC 275

    192 rThrGlnPheMetThrGln 198
        ::::   |||   :::|||
    276 TGTCTACTTCAATGCTCAG 294

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3078

seq_documentation_block:
; Sequence 3078, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 3078
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-3078


alignment_scores:
           Quality:   96.50            Length:    111
             Ratio:  1.636              Gaps:      4
 Percent Similarity:  53.153   Percent Identity: 27.027

alignment_block:
 US-09-215-435-225 x US-60-325-448-3078   ..

 Align seg 1/1  to: US-60-325-448-3078  from: 1  to: 627

    112 HisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIl 128
        ||||||:::||||||||||||   |||:::
      6 CACTGGATTGTTACTGATATACCTGGGACA.........ACAGATGCTTC 46

    128 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisS 145
          |||:::|||:::   :::|||:::::::|||:::|||      :::
     47 TTTTGGACGGGAAGTCATAAGCTATGAGAGCCCAAAGCCC......AACA 90

    145 erGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysVal 161
          |||:::||||||:::   |||   ::::::   |||:::   :::
     91 TAGGCATCCACAGGTTCATTTTTGTGCTCTTCAAGCAGAAGCGAAGGCAG 140

    162 IleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetAs 178
                :::|||         :::   |||      ::::::
    141 ACTGTAACTGTGCCT.........TCCTTCAGGGATCATTTCAACACCCG 181

    178 pArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnP 195
         :::|||      :::      ||||||   |||   |||::::::   |
    182 TCAGTTTGCTGCGGAGAATGACCTCGGCCTCCCTGTGGCAGCTGTCTACT 231

    195 heMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaProArgGlu 211
        ||      |||                                 ||||||
    232 TCAACTGTCAG.............................AGAGAG 248

    212 ArgAlaSerGluProLysHisLysAsnGlnAla 222
           |||::::::   :::   |||:::::::::
    249 ACTGCTGCCAGGAGGCGCTGAAAATCCAGCTCT 281

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4132

seq_documentation_block:
; Sequence 4132, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 4132
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-4132


alignment_scores:
           Quality:   92.50            Length:    159
             Ratio:  1.201              Gaps:      8
 Percent Similarity:  48.428   Percent Identity: 25.786

alignment_block:
 US-09-215-435-225 x US-60-325-448-4132   ..

 Align seg 1/1  to: US-60-325-448-4132  from: 1  to: 718

     93 ValMetValAspProAspAlaProSerArg...AlaGluPro.ArgGlnA 108
        |||   :::||||||||||:::|||:::      :::::|||       |
     28 GTTCATGCGGACCCAGATGTGCCAGGGCCCTAGGTGAATCCATACCCTTA 77

    108 rgPheTrpArgHisTrpLeuVal.ThrAspIleLysGly...AlaAspLe 123
        ||         ||||||:::||| :::|||||||   |||   :::|||
     78 GGGAGCATCTCCACTGGATCGTCTAGTGATATTCCTGGCACCACAGATGC 127

    123 uLysLysGlyLysIle..................................... 128
         :::   ||||||
    128 TTCTTTTGGTAAGTCCTTTTCATGCATGACTGCTTTATCTTTTGCTGATT 177

    129 .....................................GlnGlyGln 131
                                            |||:::
    178 GGTCCCATCCTTAATTTCCCTGAATAAACACATGAATGCATGCAGGAAGG 227

    132 GluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHi 148
        |||:::   :::|||:::::::|||:::|||      :::   |||:::||
    228 GAGGTGGTGAGCTACGAGAGCCCCAAGCCC......AACATTGGCATTCA 271

    148 sArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuL 165
        ||||:::   |||   ::::::   |||:::   :::
    272 CAGGTTCACCTTCGTGCTGTTCCAACAGAAGAAGCGGCAGGCCATGAACC 321

    165 euProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeu 181
          |||         :::|||||||      ::::::      ||||||
    322 CTCCC.........TCCACCAGGGACTACTTCAACACACGCCGCTTCGCT 362

    182 AsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGl 198
        ::::::      ||||||   |||   |||::::::   |||
    363 GACGAGAACGACCTTGGCCTCCCGGTCGCTGCCGTCTACTTC........ 404
```

```
     198 nAsnTyrGlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerG 215
                                 :::|||   ||||||   |||:::: 
     405 .......................AACGCGCAGCGGGAGACAGCGGCAC 429

     215 luProLysHisLysAsnGlnAla 222
         ::   :::   :::::::|||::: 
     430 GCCGCCGCTGACGGAAACAATCA 452
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6698

```
seq_documentation_block:
; Sequence 6698, Application US/09815242
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6698
;   LENGTH: 951
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(951)
US-09-815-242-6698


alignment_scores:
           Quality:   80.50            Length:    218
             Ratio:  0.782              Gaps:      9
 Percent Similarity:  47.248   Percent Identity: 20.642

alignment_block:
 US-09-215-435-225 x US-09-815-242-6698   ..

 Align seg 1/1  to: US-09-815-242-6698  from: 1  to: 951

      33 GluAlaLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPh 49
         ||||||||||||||||||||||||:::      ::::::   :::   :: 
     358 GAAGCCTTACTAGACGAAGACACTGTTGTAATCGATGCTCGTAACGATTA 407

      49 eTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnA 66
         :      :::|||||:::::         |||   ||||||   :::: 
     408 TGAATATGATTTAGGTCATTTCCGTGGTGCCGTGCGCCCAGATATCCGTA 457

      66 snTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPhePro 82
         :::::|||:::   :::   :::|||:::::: 
     458 GCTTCCGTGAG...TTACCACAATGGATTCGC..................... 486
```

```
      83 GlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAspAl 99

     486 ..................................................... 486

      99 aProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValT 116
                    |||   :::::::::|||             :::|||| 
     487 .............GAGAACAAAGAACAATTTATGGATAAAAAAATTGTTA 523

     116 hrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu 132
         ||         |||:::   ::::::      |||:::::|||
     524 CCTATTGTACTGGCGGG...ATTCGCTGTGAAAAATTTTCTGGCTGGTTA 570

     133 LeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisAr 149
         |||:::                          |||:::|||:::   :: 
     571 TTAAAAGAAGGATTTGAAGATGTTGCTCAATTGCATGGTGGTATCGCCAA 620

     149 gTyr............GlnPhePheValTyrLeuGlnGluGlyLysVal. 161
         :|||            :::           |||   :::|||||:::
     621 CTATGGGAAAAATCCAGAAACACGTGGCGAACTTTGGGACGGCAAAATGT 670

     162 ......................IleSerLeuLeuProLysGluAsnLys 170
                                :::::::::         ::::::||| 
     671 ATGTCTTTGATGACCGAATCAGTGTCGAAATTAATCATGTTGATAAAAAA 720

     171 ThrArgGly...SerTrp............LysMetAspArgPheLeuAs 182
         :::   |||   :::|||                  :::|||::::::|| 
     721 GTTATTGGGAAAGACTGGTTTGATGGAACACCTTGCGAACGCTACATTAA 770

     182 nArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnA 199
         |            ::::::|||||||::::::   |||:::::::|||:::: 
     771 C.........TGTGCAAACCCAGAATGTAATCGTCAAATCTTAACTTCAG 811

     199 snTyrGlnAspSerProThrLeu.......................Gln 207
         ::   ::::::         |||                        ||| 
     812 AAGAAAATGAACATAAACATTTAGGTGGCTGCTCATTAGAATGTAGCCAG 861

     208 AlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaGluIl 224
            |||   :::|||      :::      :::   ::::::|||||::
     862 CATCCTGCCAACCGTTATGTAAAAAAATATAATTTAACAGAAGCAGAAGT 911

     224 eAla 225
         :|||
     912 TGCT 915
```

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-6710

```
seq_documentation_block:
; Sequence 6710, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6710
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-340-6710


alignment_scores:
           Quality:   79.00            Length:     47
             Ratio:  3.038              Gaps:       1
 Percent Similarity:  55.319   Percent Identity: 36.170
```

```
alignment_block:
 US-09-215-435-225 x US-09-922-340-6710   ..

 Align seg 1/1  to: US-09-922-340-6710  from: 1  to: 439

     112 HisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIl 128
         |||||||||:::|||:::|||   |||      :::
      12 CACTGGCTGCTAACCAACATCCCGGGTAACCGGGTG...........GC 49

     128 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisS 145
         :::||||||           |||      |||   |||         |
      50 TGAAGGACAGGTGACGTGTCCCTACCTCCCCCCCTTCCCTGCCCGAGGCT 99

     145 erGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGlu 158
         ||||||:::||||||     |||:::::::::   |||:::
     100 CCGGCATCCACCGTCTTGCCTTCCTGCTCTTCAAGCAGGAC 140

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-481

seq_documentation_block:
; Sequence 481, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 481
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20774)..(21481)
US-09-897-516-481


alignment_scores:
           Quality:   79.00            Length:    201
             Ratio:  0.878              Gaps:      9
 Percent Similarity:  44.776   Percent Identity: 20.398

alignment_block:
 US-09-215-435-225 x US-09-897-516-481/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-481  from: 1  to: 24961

      18 MetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAl 34
         |||::::::|||:::   |||   :::|||:::
   22785 ATGATAATCACTTCAACTGAATTTAAGAATAAT.............. 22753

      34 aLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyrP 51
                    |||   |||         :::   ||| '    :::
   22752 .............GATTTCCTG.........AGTAAAGAGCAGGAATTCA 22725

      51 roGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyr 67
           |||:::|||:::   |||   :::      |||:::
   22724 ATGAGTTTGGTGGTAACGGTGGTAATCGCTCACCTGAA............ 22687

      68 ArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAl 84
                     :::|||   :::                  :::
   22686 ............CTTGCCTGGACAGAT.................GCGCC 22667

      84 aValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProS 101
          :::         ::::::   :::::::::   |||||||||||||||:
   22666 TGCCGGAACAAAAAGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAA 22617

     101 erArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAsp 117
         ::   :::                |||||   ||||||::::::   |||
   22616 CGGGCAGT............GGTTTCTGG...CATTGGGTGGCATTTGAT 22582

     118 Ile.....................................LysGl 120
         |||                                        ||
   22581 ATTCCGCTGGAAATACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGG 22532

     120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu.........LeuS 134
         |:::   |||      |||   ||||||:::::::::
   22531 CAGCAAATTACCGGCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGT 22482

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
           ::::::   :::      ||||||      :::      |||||||||
   22481 TTGGTTTCGGTGGAGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTAT 22432

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
           |||   |||:::   ::::::   :::   :::::::::
   22431 ATTTTCACGGTACATGCCATGTCGGTAGAAAAACTGGGTATT........ 22390

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                       :::   ::::::   :::   ::::::||||
   22389 ..................AACGCTGAAATAACAAATGCGGTCACACGTT 22359

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||   :::         :::   :::|||      :::|||      |||
   22358 TTATGATCTACGCCAATAGTCTGGCAACCGCCACTATTACAGGCTATTAC 22309

     201 Gln 201
         |||
   22308 CAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3027

seq_documentation_block:
; Sequence 3027, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3027
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (338)..(1009)
US-09-897-516-3027


alignment_scores:
           Quality:   79.00            Length:    201
             Ratio:  0.878              Gaps:      9
 Percent Similarity:  44.776   Percent Identity: 20.398

alignment_block:
```

```
US-09-215-435-225 x US-09-897-516-3027/rev   ..

Align seg 1/1  to reverse of: US-09-897-516-3027  from: 1  to: 24961

     18 MetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAl 34
        |||::::::|||:::   |||   :::|||:::
  22785 ATGATAATCACTTCAACTGAATTTAAGAATAAT..................... 22753

     34 aLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyrP 51
                   |||   |||         :::   |||      :::
  22752 ............GATTTCCTG.........AGTAAAGAGCAGGAATTCA 22725

     51 roGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyr 67
          |||:::|||:::   |||   :::      |||:::
  22724 ATGAGTTTGGTGGTAACGGTGGTAATCGCTCACCTGAA............ 22687

     68 ArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAl 84
                     :::|||   :::                 :::
  22686 ............CTTGCCTGGACAGAT.................GCGCC 22667

     84 aValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProS 101
         :::        ::::::   :::::::::   |||||||||||||||:
  22666 TGCCGGAACAAAAAGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAA 22617

    101 erArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAsp 117
        ::   :::              ||||||   ||||||::::::   |||
  22616 CGGGCAGT............GGTTTCTGG...CATTGGGTGGCATTTGAT 22582

    118 Ile.....................................LysGl 120
        |||                                      ||
  22581 ATTCCGCTGGAAATACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGG 22532

    120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu.........LeuS 134
        |:::   |||      |||   ||||||:::::::::
  22531 CAGCAAATTACCGGCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGT 22482

    134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
          ::::::   :::      ||||||      :::      |||||||||
  22481 TTGGTTTCGGTGGAGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTAT 22432

    151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
           |||   |||:::   ::::::   :::   :::::::::
  22431 ATTTTCACGGTACATGCCATGTCGGTAGAAAAACTGGGTATT........ 22390

    167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                        :::   ::::::   :::   :::::::||||
  22389 ..................AACGCTGAAATAACAAATGCGGTCACACGTT 22359

    184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
        ||   :::          :::   :::|||      :::|||      |||
  22358 TTATGATCTACGCCAATAGTCTGGCAACCGCCACTATTACAGGCTATTAC 22309

    201 Gln 201
        |||
  22308 CAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3028

seq_documentation_block:
; Sequence 3028, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3028
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2388)..(3563)
US-09-897-516-3028


alignment_scores:
           Quality:   79.00              Length:    201
             Ratio:  0.878                Gaps:      9
 Percent Similarity:  44.776   Percent Identity: 20.398

alignment_block:
 US-09-215-435-225 x US-09-897-516-3028/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3028  from: 1  to: 24961

      18 MetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAl 34
         |||:::::::|||:::   |||   :::|||:::
   22785 ATGATAATCACTTCAACTGAATTTAAGAATAAT..................... 22753

      34 aLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyrP 51
                    |||   |||         :::   |||      :::
   22752 ............GATTTCCTG.........AGTAAAGAGCAGGAATTCA 22725

      51 roGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyr 67
           |||:::|||:::   |||   :::      |||:::
   22724 ATGAGTTTGGTGGTAACGGTGGTAATCGCTCACCTGAA............ 22687

      68 ArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAl 84
                      :::|||   :::                 :::
   22686 ............CTTGCCTGGACAGAT.................GCGCC 22667

      84 aValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProS 101
          :::        ::::::   :::::::::   |||||||||||||||:
   22666 TGCCGGAACAAAAAGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAA 22617

     101 erArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAsp 117
         ::   :::              ||||||   ||||||::::::   |||
   22616 CGGGCAGT............GGTTTCTGG...CATTGGGTGGCATTTGAT 22582

     118 Ile.....................................LysGl 120
         |||                                      ||
   22581 ATTCCGCTGGAAATACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGG 22532

     120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu.........LeuS 134
         |:::   |||      |||   ||||||:::::::::
   22531 CAGCAAATTACCGGCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGT 22482

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
           ::::::   :::      ||||||      :::      |||||||||
   22481 TTGGTTTCGGTGGAGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTAT 22432

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
            |||   |||:::   ::::::   :::   :::::::::
   22431 ATTTTCACGGTACATGCCATGTCGGTAGAAAAACTGGGTATT........ 22390

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                         :::   ::::::   :::   :::::::||||
   22389 ..................AACGCTGAAATAACAAATGCGGTCACACGTT 22359

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||   :::          :::   :::|||      :::|||      |||
   22358 TTATGATCTACGCCAATAGTCTGGCAACCGCCACTATTACAGGCTATTAC 22309

     201 Gln 201
         |||
```

```
   22308 CAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3029

seq_documentation_block:
; Sequence 3029, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3029
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3565)..(3777)
US-09-897-516-3029


alignment_scores:
            Quality:   79.00              Length:    201
              Ratio:  0.878                Gaps:      9
 Percent Similarity:  44.776   Percent Identity: 20.398

alignment_block:
 US-09-215-435-225 x US-09-897-516-3029/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3029  from: 1  to: 24961

      18 MetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAl 34
         |||:::::::|||:::   |||   :::|||:::
   22785 ATGATAATCACTTCAACTGAATTTAAGAATAAT..................  22753

      34 aLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyrP 51
                    |||   |||         :::   |||      :::
   22752 .............GATTTCCTG.........AGTAAAGAGCAGGAATTCA 22725

      51 roGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyr 67
         |||:::|||:::   |||   :::      |||:::
   22724 ATGAGTTTGGTGGTAACGGTGGTAATCGCTCACCTGAA............. 22687

      68 ArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAl 84
                     :::|||   :::                      :::
   22686 .............CTTGCCTGGACAGAT.................GCGCC 22667

      84 aValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProS 101
          :::         ::::::   :::::::::   |||||||||||||||:
   22666 TGCCGGAACAAAAAGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAA 22617

     101 erArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAsp 117
         ::   :::                 ||||||   ||||||::::::   |||
   22616 CGGGCAGT............GGTTTCTGG...CATTGGGTGGCATTTGAT 22582

     118 Ile.....................................LysGl 120
         |||                                      ||
   22581 ATTCCGCTGGAAATACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGG 22532

     120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu.........LeuS 134
         |:::   |||       |||   ||||||::::::::
   22531 CAGCAAATTACCGGCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGT 22482

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ::::::   :::         ||||||       :::       |||||||||
   22481 TTGGTTTCGGTGGAGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTAT 22432

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
           |||   |||:::   ::::::   :::   :::::::::
   22431 ATTTTCACGGTACATGCCATGTCGGTAGAAAAACTGGGTATT........ 22390

     167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                           :::   ::::::   :::   ::::::||||
   22389 ..................AACGCTGAAATAACAAATGCGGTCACACGTT 22359

     184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
         ||   :::         :::   :::|||      :::|||      |||
   22358 TTATGATCTACGCCAATAGTCTGGCAACCGCCACTATTACAGGCTATTAC 22309

     201 Gln 201
         |||
   22308 CAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3030

seq_documentation_block:
; Sequence 3030, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3030
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3774)..(5156)
US-09-897-516-3030


alignment_scores:
            Quality:   79.00              Length:    201
              Ratio:  0.878                Gaps:      9
 Percent Similarity:  44.776   Percent Identity: 20.398

alignment_block:
 US-09-215-435-225 x US-09-897-516-3030/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3030  from: 1  to: 24961

      18 MetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluAl 34
         |||:::::::|||:::   |||   :::|||:::
   22785 ATGATAATCACTTCAACTGAATTTAAGAATAAT..................  22753

      34 aLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyrP 51
                    |||   |||         :::   |||      :::
   22752 .............GATTTCCTG.........AGTAAAGAGCAGGAATTCA 22725

      51 roGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyr 67
         |||:::|||:::   |||   :::      |||:::
   22724 ATGAGTTTGGTGGTAACGGTGGTAATCGCTCACCTGAA............. 22687
```

```
   68 ArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAl 84
                    :::|||   :::                  :::
22686 ............CTTGCCTGGACAGAT.................GCGCC 22667

   84 aValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProS 101
       :::         ::::::   :::::::::   ||||||||||||||:
22666 TGCCGGAACAAAAAGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAA 22617

  101 erArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAsp 117
      ::   :::                |||||   |||||::::::   |||
22616 CGGGCAGT............GGTTTCTGG...CATTGGGTGGCATTTGAT 22582

  118 Ile.....................................LysGl 120
      |||                                     ||
22581 ATTCCGCTGGAAATACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGG 22532

  120 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu.........LeuS 134
      |:::   |||      |||   ||||||:::::::::
22531 CAGCAAATTACCGGCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGT 22482

  134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
       ::::::   :::      ||||||      :::      |||||||||
22481 TTGGTTTCGGTGGAGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTAT 22432

  151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         |||   |||:::   ::::::   :::   :::::::::
22431 ATTTTCACGGTACATGCCATGTCGGTAGAAAAACTGGGTATT........ 22390

  167 sGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgP 184
                      :::   ::::::   :::   ::::::||||
22389 ..................AACGCTGAAATAACAAATGCGGTCACACGTT 22359

  184 heHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyr 200
      ||   :::         :::   :::|||      :::|||      |||
22358 TTATGATCTACGCCAATAGTCTGGCAACCGCCACTATTACAGGCTATTAC 22309

  201 Gln 201
      |||
22308 CAG 22306
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:31 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225
Query length: 227
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est102:BG545668     +  1028.50 2089.36  2.9e-107  749  ! BG545668 602572933F1 NIH_MGC_77
gb_est93:BF854481      +   921.00 1871.93  3.8e-95   564  ! BF854481 MR2-EN0093-271200-005-
gb_est93:BF852989      +   884.00 1795.26  7.1e-91   603  ! BF852989 MR2-EN0093-191200-001-
gb_est93:BF852995      +   853.00 1732.41  2.2e-87   563  ! BF852995 MR2-EN0093-191200-001-
gb_est93:BF853000      +   836.00 1697.63  1.9e-85   558  ! BF853000 MR2-EN0093-191200-001-
gb_est93:BF853920      +   792.00 1608.59  1.8e-80   499  ! BF853920 MR2-EN0093-261200-004-
gb_est93:BF853928      +   785.00 1593.42  1.2e-79   536  ! BF853928 MR2-EN0093-261200-004-
gb_est93:BF846063      -   763.00 1549.89  3.3e-77   464  ! BF846063 QV0-EN0055-181000-439-
gb_est93:BF848191      +   735.50 1493.48  4.5e-74   463  ! BF848191 QV0-EN0055-031100-480-
gb_est93:BF846065      -   725.50 1472.87  6.4e-73   467  ! BF846065 QV0-EN0055-181000-439-
gb_est93:BF848190      +   723.50 1468.86  1.1e-72   463  ! BF848190 QV0-EN0055-031100-480-
gb_est25:AI816715      -   718.00 1455.71  5.8e-72   547  ! AI816715 wj43b08.x1 NCI_CGAP_Lu
gb_est25:AI857902      -   718.00 1455.71  5.8e-72   547  ! AI857902 wj68a01.x1 NCI_CGAP_Lu
gb_est69:BE044451      -   709.00 1437.75  5.8e-71   523  ! BE044451 ho45g08.x1 Soares_NFL_
gb_est93:BF846059      +   705.00 1430.92  1.4e-70   462  ! BF846059 QV0-EN0055-181000-439-
gb_est17:AI218954      -   699.00 1416.62  8.7e-70   552  ! AI218954 qg72g08.x1 Soares_NFL_
gb_est26:AI936064      -   695.00 1408.79  2.4e-69   534  ! AI936064 wo47h12.x1 NCI_CGAP_Ga
gb_est93:BF846558      +   687.50 1393.44  1.7e-68   532  ! BF846558 QV0-EN0058-261000-441-
gb_est110:W32197       +   686.00 1392.20  2.0e-68   451  ! W32197 zc04f05.r1 Soares_parath
gb_est10:AA688029      -   675.00 1368.24  4.3e-67   511  ! AA688029 nv16e12.s1 NCI_CGAP_Pr
gb_est93:BF848198      +   672.50 1364.21  7.2e-67   463  ! BF848198 QV0-EN0055-031100-480-
gb_est3:AA192426       +   627.00 1271.66  1.0e-61   430  ! AA192426 zq01g08.r1 Stratagene
gb_est110:W37255       +   626.00 1269.06  1.4e-61   452  ! W37255 zc09f01.r1 Soares_parath
gb_est101:BG483879     +   611.50 1233.09  1.4e-59   789  ! BG483879 602504455F1 NIH_MGC_77
gb_est104:F20527       +   605.00 1227.46  3.0e-59   395  ! F20527 HSPD04880 HM3 Homo sapie
gb_est104:F28028       +   604.00 1224.65  4.3e-59   423  ! F28028 HSPD16404 HM3 Homo sapie
gb_est53:AW975331      -   582.00 1178.60  1.6e-56   459  ! AW975331 EST387439 MAGE reseque
gb_est104:F24489       +   581.00 1180.09  1.3e-56   334  ! F24489 HSPD10829 HM3 Homo sapie
gb_est93:BF853537      +   560.00 1134.79  4.3e-54   407  ! BF853537 MR2-EN0093-211200-002-
gb_est3:AA192427       -   556.00 1125.29  1.5e-53   457  ! AA192427 zq01g08.s1 Stratagene
gb_htc:AK006964        +   555.00 1113.99  6.2e-53  1048  ! AK006964 Mus musculus adult ma
gb_est90:BF599823      +   540.50 1094.61  7.5e-52   413  ! BF599823 264147 MARC 3BOV Bos t
gb_est104:F33888       +   529.00 1071.32  1.5e-50   402  ! F33888 HSPD28217 HM3 Homo sapie
gb_est93:BF890431      +   521.00 1053.55  1.5e-49   454  ! BF890431 292000 MARC 3BOV Bos t
gb_est13:AA903899      -   491.50  993.36  3.3e-46   440  ! AA903899 oe78e07.s1 NCI_CGAP_Lu
gb_est104:F26660       +   490.00  995.32  2.5e-46   280  ! F26660 HSPD14219 HM3 Homo sapie
gb_est12:AA854779      -   487.00  984.86  9.7e-46   412  ! AA854779 aj77b05.s1 Soares_para
gb_est105:H83784       +   482.00  973.87  4.0e-45   440  ! H83784 ys64b02.r1 Soares retina
gb_est104:F19507       +   481.00  976.97  2.7e-45   277  ! F19507 HSPD04111 HM1 Homo sapie
gb_est104:F36903       +   479.00  967.81  8.7e-45   436  ! F36903 HSPD34824 HM3 Homo sapie
gb_est10:AA661735      -   474.00  958.51  2.9e-44   400  ! AA661735 nv04b11.s1 NCI_CGAP_Pr
gb_est71:BE165776      +   449.00  907.18  2.1e-41   401  ! BE165776 CM3-HT0486-310100-0
gb_est105:H83927       -   444.00  898.38  6.4e-41   352  ! H83927 ys64b02.s1 Soares ret
gb_est110:W33189       -   444.00  895.97  8.7e-41   437  ! W33189 zc09f01.s1 Soares_par
gb_est104:F31628       +   436.00  885.42  3.4e-40   258  ! F31628 HSPD22948 HM3 Homo sa

seq_name: gb_est102:BG545668

seq_documentation_block:
LOCUS       BG545668      749 bp    mRNA            EST       04-APR-2001
DEFINITION  602572933F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4700840 5',
            mRNA sequence.
ACCESSION   BG545668
VERSION     BG545668.1  GI:13544333
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1535  row: b  column: 09
            High quality sequence stop: 740.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4700840"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      184 a    206 c    205 g    154 t
ORIGIN


alignment_scores:
            Quality: 1028.50            Length:    224
              Ratio:  4.851               Gaps:      5
 Percent Similarity:  94.643   Percent Identity: 93.304

alignment_block:
 US-09-215-435-225 x BG545668   ..

 Align seg 1/1  to: BG545668  from: 1  to: 749

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      86 ATGGGTTGGACAATGAGGCTGGTTACAGCAGCACTGTTACTGGGTCTCAT 135

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     136 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 185

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
```

```
     |||||||||||||||||||||||||||||   ||||||||||||||||||
   186 CCCTCTTGGACGAGGACACCCTCTTTTGCCA.GGCCTTGAAGTTTTCTAC 234

    51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   235 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 284

    67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   285 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 334

    84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   335 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 384

   101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   385 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 434

   117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
     |||||||||||||||||||||:::|||||||||||||||||||||||||||
   435 TATCAAGGGCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTAT 484

   134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
     |||||||||||||||||||||||   |||||||||||||||||||||||||
   485 CAGCCTACCAGGCTCCCTCCCCAACGGCACACAGTGGCTTCCATCGCTAC 534

   151 GlnPhePheValTyrLeuGlnGluGly.LysValIleSerLeuLeuProL 167
     |||||||||||||||||||||||||||| :::|||||||||||||||||||
   535 CAGTTCTTTGTCTATCTTCAGGAAGGAACAAGTCATCTCTCTCCTTCCCA 584

   167 ysGluAsnLysThrArgGlySerTrpLysMetAspArg...PheLeuAsn 182
     |||||||||||||||||||||||||:::            |||||||||
   585 AGGAAAACAAAACTCGAGGCTCTTGGAACAATGGCACAGATTTCTTGAAC 634

   183 ArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMet.ThrGlnA 199
     |||||||||||||||||||||||||||||||||||||||| |||||||
   635 CGGTTCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGAACCCAGA 684

   199 sn.TyrGlnAspSer.ProThrLeuGlnAlaProArgGluArgAlaSerG 215
     || |||||||||||| |||||||||||||||||||   ||||||||||
   685 ACTTACCAGGACTCACCCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCG 734

   215 luProLysHisLys 219
     |||||      |||
   735 AGCCAGCACCAAAA 748

seq_name: gb_est93:BF854481

seq_documentation_block:
LOCUS       BF854481      564 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-271200-005-d12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF854481
VERSION     BF854481.1  GI:12242225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            271200-005-d12&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      149 a    156 c    141 g    118 t
ORIGIN


alignment_scores:
            Quality:  921.00              Length:    183
              Ratio:  5.174                 Gaps:      1
 Percent Similarity:  97.268   Percent Identity: 94.536

alignment_block:
 US-09-215-435-225 x BF854481   ..

 Align seg 1/1  to: BF854481  from: 1  to: 564

    17 MetMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGl 33
       :::::::|||:::|||      :::   :::|||||||||||||||||||
    16 CTGTTTGTGACCACCATCATGAGACCCAGTAACAGCCCGTGTGCCCATGA 65

    33 uAlaLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheT 50
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    66 GGCCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCT 115

    50 yrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsn 66
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   116 ACCCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAAC 165

    67 TyrArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGl 83
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   166 TACAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGG 215

    83 yAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAspAla. 99
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   216 GGCCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCC 265

   100 ProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValTh 116
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   266 CCTAGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAAC 315

   116 rAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluL 133
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   316 AGATATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGT 365

   133 euSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArg 149
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   366 TATCAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGC 415

   150 TyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuPr 166
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   416 TACCAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCC 465
```

```
     166 oLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsnA 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     466 CAAGGAAAACAAAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACC 515

     183 rgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGln 198
         ||||||||||||||||   ||||||||||||||||||||||||||
     516 GTTTCCACCTGGGCGAAACTGAAGCAAGCACCCAGTTCATGACCCAG 562

seq_name: gb_est93:BF852989

seq_documentation_block:
LOCUS       BF852989      603 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-a12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852989
VERSION     BF852989.1  GI:12240837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-a12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 560.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      140 a    157 c    171 g    135 t
ORIGIN


alignment_scores:
            Quality:  884.00            Length:    171
              Ratio:  5.293               Gaps:      0
 Percent Similarity:  97.661   Percent Identity: 97.076

alignment_block:
 US-09-215-435-225 x BF852989   ..

 Align seg 1/1  to: BF852989  from: 1  to: 603
```

```
       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 190

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||:::||||||||||||||||||||||||||||||||||||||
     191 CCCTCTTGAACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 240

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 290

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 340

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 390

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 440

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGGCAGGAGTTAT 490

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAA 540

     151 GlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLy 167
         ||||||||||||||||||||   ||||||||||||||||||||   ||
     541 CAGTTCTTTGTCTATCTTCAGGGAGGGAAAGTCATCTCTCTCCTTTCCAA 590

     167 sGluAsnLysThr 171
         |   |||||||||
     591 GGGAAACAAAACT 603

seq_name: gb_est93:BF852995

seq_documentation_block:
LOCUS       BF852995      563 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-e01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852995
VERSION     BF852995.1  GI:12240843
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
```

```
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
          191200-001-e01&t3=2000-12-19&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      126 a    148 c    164 g    125 t
ORIGIN


alignment_scores:
            Quality:  853.00            Length:    157
              Ratio:  5.433               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x BF852995   ..

 Align seg 1/1  to: BF852995  from: 1  to: 563

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 190

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 240

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 290

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 340

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 390

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 440

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 490

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 CAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTAC 540

     151 GlnPhePheValTyrLeuGln 157
         |||||||||||||||||||||
     541 CAGTTCTTTGTCTATCTTCAG 561

seq_name: gb_est93:BF853000

seq_documentation_block:
LOCUS       BF853000      558 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-f12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853000
VERSION     BF853000.1  GI:12240848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-f12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 558.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      124 a    146 c    163 g    124 t      1 others
ORIGIN


alignment_scores:
            Quality:  836.00            Length:    155
              Ratio:  5.429               Gaps:      0
 Percent Similarity:  99.355   Percent Identity: 99.355

alignment_block:
 US-09-215-435-225 x BF853000   ..

 Align seg 1/1  to: BF853000  from: 1  to: 558

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      93 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 142

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     143 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 192
```

```
      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     193 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 242

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     243 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 292

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     293 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 342

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     343 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 392

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     393 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 442

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     443 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGNCAGGAGTTAT 492

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyr 150
         ||||||||||||||||||||||   ||||||||||||||||||||||||||
     493 CAGCCTACCAGGCTCCCTCCCCAGCGGCACACAGTGGCTTCCATCGCTAC 542

     151 GlnPhePheValTyr 155
         ||||||||||||||||
     543 CAGTTCTTTGTCTAT 557
```

seq_name: gb_est93:BF853920

seq_documentation_block:

```
LOCUS       BF853920      499 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-d01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853920
VERSION     BF853920.1  GI:12241664
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            261200-004-d01&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 227.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      119 a    129 c    148 g    103 t
ORIGIN
```

alignment_scores:

```
           Quality:  792.00             Length:    149
             Ratio:  5.351                Gaps:      0
 Percent Similarity:  99.329   Percent Identity: 97.987
```

alignment_block:

US-09-215-435-225 x BF853920 ..

Align seg 1/1 to: BF853920 from: 1 to: 499

```
       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      48 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 97

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      98 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 147

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     148 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 197

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     198 CCAGAGTTGGGGAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTA 247

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||   |||||||
     248 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTCCCCGGGGG 297

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||:::||||||||||||||||||||||||||||||||
     298 CCGTGGACGGCGCAAACTATATCCTGGTGATGGTGGATCCAGATGCCCCT 347

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     348 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 397

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     398 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTAT 447

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArg 149
         ||:::||||||||||||||||||||||||||||||||||||||||||
     448 CAGGCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGC 494
```

seq_name: gb_est93:BF853928

seq_documentation_block:

```
LOCUS       BF853928      536 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-g12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853928
VERSION     BF853928.1  GI:12241672
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brèntani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            261200-004-g12&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      121 a    139 c    164 g    112 t
ORIGIN


alignment_scores:
            Quality:  785.00            Length:    148
              Ratio:  5.340               Gaps:      0
 Percent Similarity:  99.324   Percent Identity: 97.973

alignment_block:
 US-09-215-435-225 x BF853928   ..

 Align seg 1/1  to: BF853928  from: 1  to: 536

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGlyAspGluAspGluAsnSerProCysAlaHisGluA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 GATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGG 190

      34 laLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPheTyr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 CCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTAC 240

      51 ProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 CCAGAGTTGGGGAACATTGGCTGCAAGGTCGTTCCTGATTGTAACAACTA 290

      67 rArgGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 CAGACAGAAGATCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGG 340

      84 laValAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCGTGGACGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCT 390

     101 SerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 AGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGA 440

     117 pIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuS 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAGGGGCAGGAGTTAT 490

     134 erAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHis 148
         ||:::  |||:::|||||||||||||||||||||||||||||||
     491 CAGGCTAACAGGGTCCCTCCCCACCGGCACACAGTGGCTTCCAT 534

seq_name: gb_est93:BF846063

seq_documentation_block:
LOCUS       BF846063      464 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-181000-439-f11 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846063
VERSION     BF846063.1  GI:12233213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            181000-439-f11&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       94 a    134 c    122 g    114 t
ORIGIN


alignment_scores:
            Quality:  763.00            Length:    142
              Ratio:  5.373               Gaps:      0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225 x BF846063/rev   ..

 Align seg 1/1  to reverse of: BF846063  from: 1  to: 464

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     444 ATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCATGATGGTGGTCAC 395

      21 rGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspG 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     394 TGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACG 345

      38 luAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGly 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     344 AGGACACCCTCTTTTGCCAGGGCCTTGAAGTCTTCTACCCAGAGTTGGGG 295

      55 AsnIleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIl 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     294 AACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGAT 245

      71 eThrSerTrpMetGluProIleValLysPheProGlyAlaValAspGlyA 88
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     244 CACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCG 195

      88 laThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGlu 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     194 CAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAA 145

     105 ProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAl 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     144 CCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGC 95

     121 aAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnA 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      94 CGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGG 45

     138 laProSerProProAlaHisSerGly 146
         |||||||||||||||||||||||||
      44 CTCCCTCCCCACCGGCACACAGTGGC 19

seq_name: gb_est93:BF848191

seq_documentation_block:
LOCUS       BF848191      463 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-031100-480-c10 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF848191
VERSION     BF848191.1  GI:12235341
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            031100-480-c10&t3=2000-11-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      116 a    119 c    136 g     92 t
ORIGIN


alignment_scores:
            Quality:  735.50             Length:    145
              Ratio:  5.180                Gaps:      1
 Percent Similarity:  97.931   Percent Identity: 95.172

alignment_block:
 US-09-215-435-225 x BF848191   ..

 Align seg 1/1  to: BF848191  from: 1  to: 463

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||   :::||||||||||||||||||||||||
      30 ATGAGGCTGGTCACAGCAGCA...GTACTGGGTCTCATGATGGTGGTCAC 76

      21 rGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspG 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      77 TGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACG 126

      38 luAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGly 54
         ||||||||||||||||||||||   ||||||||||||||||||||||||
     127 AGGACACCCTCTTTTGCCAGGGCCATGAAGTTTTCTACCCAGAGTTGGGG 176

      55 AsnIleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIl 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     177 AACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGAT 226

      71 eThrSerTrpMetGluProIleValLysPheProGlyAlaValAspGlyA 88
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     227 CACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCG 276

      88 laThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGlu 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     277 CAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAA 326

     105 ProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAl 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     327 CCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGC 376

     121 aAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnA 138
         ||||||||||:::||||||||||||||||||||||||||||||||||||
     377 CGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGG 426

     138 laProSerProProAlaHisSerGlyPheHisArg 149
         |||||||||||||||||||||||:::   :::
     427 CTCCCTCCCCACCGGCACACAGTGGCATCGATAAA 461

seq_name: gb_est93:BF846065
```

```
seq_documentation_block:
LOCUS       BF846065      467 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-181000-439-g06 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846065
VERSION     BF846065.1  GI:12233215
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            181000-439-g06&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 467.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       94 a    134 c    126 g    113 t
ORIGIN


alignment_scores:
           Quality:  725.50            Length:    143
             Ratio:  5.182               Gaps:      2
 Percent Similarity:  97.902   Percent Identity: 97.203

alignment_block:
 US-09-215-435-225 x BF846065/rev   ..

 Align seg 1/1  to reverse of: BF846065  from: 1  to: 467

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     447 ATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCATGATGGTGGTCAC 398

      21 rGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspG 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     397 TGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACG 348

      38 luAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGly 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     347 AGGACACCCTCTTTTGCCAGGGCCTTGAAGTCTTCTACCCAGAGTTGGGG 298

      55 AsnIleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIl 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     297 AACATCGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGAT 248

      71 eThrSerTrpMetGluProIleValLysPheProGlyAlaValAspGlyA 88
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     247 CACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCG 198

      88 laThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGlu 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     197 CAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAA 148

     105 ProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAl 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     147 CCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGC 98

     121 aAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnA 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      97 CGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC...C 51

     138 la.ProSerProProAlaHisSerGly 146
          ||||||||||||||||||:::|||
      50 GCTCCCTCCCCACCGGCACACGGTGGC 24

seq_name: gb_est93:BF848190

seq_documentation_block:
LOCUS       BF848190      463 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-031100-480-c05 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF848190
VERSION     BF848190.1  GI:12235340
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            031100-480-c05&t3=2000-11-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 25
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
```

```
                No. 196,716 -  Ludwig Institute for Cancer Research)
                profiles into  the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      117 a    120 c    136 g     90 t
ORIGIN


alignment_scores:
           Quality:  723.50            Length:    146
             Ratio:  5.095               Gaps:      2
 Percent Similarity:  97.260   Percent Identity: 94.521

alignment_block:
 US-09-215-435-225 x BF848190   ..

 Align seg 1/1  to: BF848190  from: 1  to: 463

       5 MetArgLeuValThrAlaAla.LeuLeuLeuGlyLeuMetMetValValT 21
         |||||||||||||||||||| :::   ||||||||||||||||||||
      29 ATGAGGCTGGTCACAGCAGCACGTA...CTGGGTCTCATGATGGTGGTCA 75

      21 hrGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAsp 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 CTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGAC 125

      38 GluAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGl 54
         ||||||||||||||||||||   |||||||||:::|||||||||||||||
     126 GAGGACACCCTCTTTTGCCAGGACCTTGAAGTCTACTACCCAGAGTTGGG 175

      54 yAsnIleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysI 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     176 GAACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGA 225

      71 leThrSerTrpMetGluProIleValLysPheProGlyAlaValAspGly 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     226 TCACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGC 275

      88 AlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGl 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     276 GCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGA 325

     104 uProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyA 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     326 ACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCG 375

     121 laAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGln 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 CCGACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAG 425

     138 AlaProSerProProAlaHisSerGlyPheHisArg 149
         ||||||||||||||||||||||||||:::   :::
     426 GCTCCCTCCCCACCGGCGCACAGTGGCATCGATAAA 461

seq_name: gb_est25:AI816715

seq_documentation_block:
LOCUS       AI816715      547 bp    mRNA            EST       21-DEC-1999
DEFINITION  wj43b08.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405559 3'
            similar to SW:D2_ONCVO P54187 D2 PROTEIN ;, mRNA sequence.
ACCESSION   AI816715
VERSION     AI816715.1  GI:5435794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 903   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2405559"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      102 a    131 c    160 g    154 t
ORIGIN


alignment_scores:
           Quality:  718.00            Length:    141
             Ratio:  5.129               Gaps:      0
 Percent Similarity:  99.291   Percent Identity: 98.582

alignment_block:
 US-09-215-435-225 x AI816715/rev   ..

 Align seg 1/1  to reverse of: AI816715  from: 1  to: 547

      87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     547 GGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGC 498

     103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     497 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 448

     120 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 136
         |||||||||||||||:::|||||||||||||||||||||||||||||||
     447 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 398

     137 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     397 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 348

     153 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     347 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 298

     170 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     297 AAACTCGAGGCTCTTGGAAAATGGACAGATTTTTGAACCGCTTCCACTTG 248

     187 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     247 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 198
```

```
    203 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 220
        ||||||||||||||||||||||   ||||||||||||||||||||||   |
    197 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 149

    220 snGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||
    148 ACCAGGCGGAGATAGCTGCCTGC 126

seq_name: gb_est25:AI857902

seq_documentation_block:
LOCUS       AI857902      547 bp    mRNA            EST       21-DEC-1999
DEFINITION  wj68a01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407944 3'
            similar to SW:D2_ONCVO P54187 D2 PROTEIN ;, mRNA sequence.
ACCESSION   AI857902
VERSION     AI857902.1  GI:5511518
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 752   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2407944"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      102 a    130 c    160 g    155 t
ORIGIN


alignment_scores:
            Quality:  718.00            Length:    141
              Ratio:  5.129               Gaps:      0
 Percent Similarity:  99.291   Percent Identity: 98.582

alignment_block:
 US-09-215-435-225 x AI857902/rev   ..

 Align seg 1/1  to reverse of: AI857902  from: 1  to: 547

     87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    547 GGCGCAACCTATATCTTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGC 498

    103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    497 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 448

    120 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 136
        ||||||||||||||:::||||||||||||||||||||||||||||||||
    447 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 398

    137 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 153
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    397 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 348

    153 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 170
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    347 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 298

    170 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 186
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    297 AAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTG 248

    187 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 203
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    247 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 198

    203 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 220
        ||||||||||||||||||||||   ||||||||||||||||||||||   |
    197 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 149

    220 snGlnAlaGluIleAlaAlaCys 227
        |||||||||||||||||||||||
    148 ACCAGGCGGAGATAGCTGCCTGC 126

seq_name: gb_est69:BE044451

seq_documentation_block:
LOCUS       BE044451      523 bp    mRNA            EST       08-JUN-2000
DEFINITION  ho45g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:3040382 3' similar to SW:D1_ONCVO P54186 D1 PROTEIN ;, mRNA
            sequence.
ACCESSION   BE044451
VERSION     BE044451.1  GI:8361504
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3040382"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
```

```
                NCI_CGAP_GCB1) were mixed, and ss circles were made in
                vitro. Following HAP purification, this DNA was used as
                tracer in a subtractive hybridization reaction. The driver
                was PCR-amplified cDNAs from pools of 5,000 clones made
                from the same 3 libraries. The pools consisted of
                I.M.A.G.E. clones 297480-302087, 682632-687239,
                726408-728711, and 729096-731399. Subtraction by Bento
                Soares and M. Fatima Bonaldo. "
BASE COUNT       96 a    128 c    159 g    140 t
ORIGIN


alignment_scores:
            Quality:  709.00             Length:    141
              Ratio:  5.101                Gaps:      0
 Percent Similarity:  98.582   Percent Identity: 97.163

alignment_block:
 US-09-215-435-225 x BE044451/rev   ..

 Align seg 1/1  to reverse of: BE044451  from: 1  to: 523

      87 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 103
         ||||||||||||||||   :::||||||||||||||||||||||||||||||
     523 GGCGCAACCTATATCTGGCTGATGGTGGATCCAGATGCCCCTAGCAGAGC 474

     103 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     473 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 424

     120 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 136
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
     423 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 374

     137 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     373 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 324

     153 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     323 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 274

     170 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     273 AAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTG 224

     187 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     223 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 174

     203 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 220
         ||||||||||||||||||||||   ||||||||||||||||||||   |
     173 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 125

    \220 snGlnAlaGluIleAlaAlaCys 227
         |||||||||||||||||||||||
     124 ACCAGGCGGAGATAGCTGCCTGC 102

seq_name: gb_est93:BF846059

seq_documentation_block:
LOCUS       BF846059      462 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-181000-439-e09 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846059
VERSION     BF846059.1  GI:12233209
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            181000-439-e09&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      113 a    121 c    134 g     94 t
ORIGIN


alignment_scores:
            Quality:  705.00             Length:    145
              Ratio:  4.896                Gaps:      0
 Percent Similarity:  99.310   Percent Identity: 97.241

alignment_block:
 US-09-215-435-225 x BF846059   ..

 Align seg 1/1  to: BF846059  from: 1  to: 462

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||||||||||   ||||||||||||||||||||
      28 ATGAGGCTGGTCACAGCAGCACTGTTACT.GGTCTCATGATGGTGGTCAC 76

      21 rGlyAspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspG 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      77 TGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACG 126

      38 luAspThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGly 54
         ||||||||||||||||   ||||||||||||||||||||||||||||||
     127 AGGACACCCTCTTTTGCCA.GGCCTTGAAGTCTTCTACCCAGAGTTGGGG 175

      55 AsnIleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIl 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     176 AACATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGAT 225

      71 eThrSerTrpMetGluProIleValLysPheProGlyAlaValAspGlyA 88
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     226 CACCTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCG 275

      88 laThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGlu 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     276 CAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAA 325
```

```
105 ProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAl 121
    ||||||||||||||||||||||||||||||||||||||||||||||||||
326 CCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGC 375

121 aAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnA 138
    |||||||:::||||||||||||||||||||||||||||||||||||||||
376 CGACCTGCAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGG 425

138 laProSerProProAlaHisSerGlyPheHisArg 149
    ||||||||||||||||||||||||:::   :::
426 CTCCCTCCCCACCGGCGCACAGTGGCATCGATAAA 460
```

```
OM of: US-09-215-435-225_COPY_1_22 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:51 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat2:AX083412      +  109.00  325.24  7.0e-10   681  ! AX083412 Sequence 104 from Pate
gb_pat2:AX083422      +  109.00  322.98  9.4e-10   897  ! AX083422 Sequence 114 from Pate
gb_pat1:AX060293      +  109.00  322.67  9.8e-10   932  ! AX060293 Sequence 1 from Patent
gb_htg12:AC037441     -  109.00  280.29  2.2e-07  162795  ! AC037441 Homo sapiens clone
gb_htg17:AC084081     -  109.00  279.95  2.3e-07  169645  ! AC084081 Homo sapiens chromo
gb_pr2:AC005723       +   58.00  122.37  140.36   127766  ! AC005723 Homo sapiens chromo
gb_htg1:AC008437      -   58.00  120.04  189.24   169700  ! AC008437 Homo sapiens chromo
gb_htg1:AC008588      -   58.00  118.44  232.33   206211  ! AC008588 Homo sapiens chromo
gb_ro1:MUSNECDIN      -   55.00  148.99  4.62      1585  ! M80840 M.musculus necdin mRNA,
gb_ro1:MUSNCP         -   55.00  141.60  11.91     3899  ! D76440 Mouse gene for necdin,
gb_htg11:AC026388_0   -   55.00  114.19  400.62   110000  ! AC026388 Mus musculus chromo
gb_ro1:AC027298       +   55.00  109.22  757.55   201467  ! AC027298 Mus musculus 7 BAC
gb_ov:ORZHSC70        +   54.00  142.94  10.03     2261  ! D13669 Oryzias latipes mRNA fo
gb_pl2:AF193060       +   54.00  138.39  17.98     3937  ! AF193060 Phanerochaete chrysos
gb_htg4:AC014412      +   54.00  113.03  464.74   86446  ! AC014412 Drosophila melanogas
gb_htg13:AC055859     -   54.00  108.22  861.09   155295  ! AC055859 Homo sapiens chromo
gb_pr2:AC007370       +   54.00  107.17  984.86   176426  ! AC007370 Homo sapiens, clone
gb_in1:AC009213       -   54.00  106.90  1.0e+03  182525  ! AC009213 Drosophila melanoga
gb_in2:AE003765       -   54.00  104.68  1.4e+03  239171  ! AE003765 Drosophila melanoga
gb_pl2:AP002844       -   53.00  104.60  1.4e+03  164839  ! AP002844 Oryza sativa genomi
gb_pl4:VCSSG185       -   52.00  136.03  24.33     2444  ! X51616 V.carteri RNA for sulfa
gb_ro1:MMU41741       +   52.00  123.92  115.04    10690  ! U41741 Mus musculus transcrip
gb_pl3:BRRB1F25       +   51.00  146.24  6.57       481  ! L31937 Brassica rapa (clone bif
gb_ba3:SFU46133       +   51.00  134.94  27.98     1905  ! U46133 Shigella flexneri amino
gb_ba1:AE005356       +   51.00  122.12  144.88    9083  ! AE005356 Escherichia coli O157
gb_ba1:AE000246       -   51.00  117.30  268.80    16338  ! AE000246 Escherichia coli K12
gb_ba2:D90790         -   51.00  116.50  297.89    18013  ! D90790 E.coli genomic DNA, Kc
gb_ba2:D90791         -   51.00  115.67  331.34    19929  ! D90791 E.coli genomic DNA, Kc
gb_pr9:HSJ245M18      +   51.00  102.05  1.9e+03  104770  ! AL078584 Human DNA sequence
gb_pr8:HS222E13       -   51.00  100.71  2.3e+03  123360  ! Z93241 Human DNA sequence fr
gb_htg25:HS919B11     -   51.00   98.53  3.0e+03  160725  ! AL031302 Homo sapiens chromo
gb_htg20:AL157892     +   51.00   97.64  3.3e+03  179116  ! AL157892 Homo sapiens chromo
gb_htg22:AL512428     -   51.00   96.99  3.6e+03  193884  ! AL512428 Homo sapiens chromo
gb_htg22:AL392187     -   51.00   95.83  4.2e+03  223335  ! AL392187 Homo sapiens chromo
gb_ba2:AP002557       -   51.00   93.79  5.5e+03  286485  ! AP002557 Escherichia coli O1
gb_ba3:PYWATRNAS      +   50.50  135.06  27.54     1550  ! D45167 Pyrococcus sp. gene for
gb_vi2:PACAPSID       +   50.50  131.28  44.76     2459  ! Y15938 Porcine astrovirus gene
gb_vi1:AB037272       +   50.50  130.42  49.97     2730  ! AB037272 Porcine astrovirus ge
gb_ro1:AB000716S1     +   50.00  129.34  57.41     2573  ! AB000716 Rattus rattus gene fo
gb_pl4:SPCC74         +   50.00  107.76  913.39    35637  ! AL031543 S.pombe chromosome I
gb_pl1:AC005936       +   50.00  101.05  2.2e+03   80735  ! AC005936 Arabidopsis thaliana
gb_htg23:AL590149     +   50.00   94.98  4.7e+03   169181  ! AL590149 Danio rerio clon
gb_htg3:AC011676      -   50.00   93.75  5.5e+03   196574  ! AC011676 Homo sapiens clo
gb_htg2:AC009711      -   50.00   93.49  5.7e+03   202865  ! AC009711 Homo sapiens chr
gb_ro2:RATADREC1B     +   49.50  129.63  55.26      2050  ! L28752 Rat adrenergic recep

seq_name: gb_pat2:AX083412

seq_documentation_block:
LOCUS       AX083412      681 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 104 from Patent WO0112660.
ACCESSION   AX083412
VERSION     AX083412.1  GI:13185249
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 104 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      179 a    186 c    183 g    133 t
ORIGIN


alignment_scores:
           Quality:  109.00             Length:     22
             Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AX083412   ..

 Align seg 1/1  to: AX083412  from: 1  to: 681

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 50

      17 tMetValValThrGly 22
         ||||||||||||||||
      51 GATGGTGGTCACTGGA 66

seq_name: gb_pat2:AX083422

seq_documentation_block:
LOCUS       AX083422      897 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 114 from Patent WO0112660.
ACCESSION   AX083422
VERSION     AX083422.1  GI:13185262
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 114 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             99. .782
                     /note="unnamed protein product"
```

```
                     /codon_start=1
                     /protein_id="CAC33305.1"
                     /db_xref="GI:13185263"
                     /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
                     GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                     AEPRQRFWRHWLVTDIKGADLKKGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                     VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                     KNQAEIAAC"
BASE COUNT       226 a    253 c    234 g    184 t
ORIGIN


alignment_scores:
            Quality:  109.00             Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AX083422   ..

 Align seg 1/1  to: AX083422  from: 1  to: 897

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 148

       17 tMetValValThrGly 22
          ||||||||||||||||
     149 GATGGTGGTCACTGGA 164

seq_name: gb_pat1:AX060293

seq_documentation_block:
LOCUS        AX060293      932 bp    DNA             PAT       22-JAN-2001
DEFINITION   Sequence 1 from Patent WO0078802.
ACCESSION    AX060293
VERSION      AX060293.1  GI:12405782
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 932)
  AUTHORS    Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and
             Herrmann,J.L.
  TITLE      Secreted polypeptides and corresponding polynucleotides
  JOURNAL    Patent: WO 0078802-A 1 28-DEC-2000;
             Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             113. .796
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC24866.1"
                     /db_xref="GI:12405783"
                     /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
                     GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                     AEPRQRFWRHWLVTDIKGADLKEGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                     VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                     KNQAEIAAC"
BASE COUNT       244 a    257 c    244 g    187 t
ORIGIN


alignment_scores:
            Quality:  109.00             Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AX060293   ..

 Align seg 1/1  to: AX060293  from: 1  to: 932

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     113 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 162

       17 tMetValValThrGly 22
          ||||||||||||||||
     163 GATGGTGGTCACTGGA 178

seq_name: gb_htg12:AC037441

seq_documentation_block:
LOCUS        AC037441   162795 bp    DNA             HTG       03-MAR-2001
DEFINITION   Homo sapiens clone RP11-87E22, WORKING DRAFT SEQUENCE, 11 unordered
             pieces.
ACCESSION    AC037441
VERSION      AC037441.3  GI:13184260
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 162795)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome, clone RP11-87E22
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 162795)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
             Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
             Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
             Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
             Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
             Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
             Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
             Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
             Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
             McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
             Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
             Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
             O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
             Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
             Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
             Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
             Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Mar 2, 2001 this sequence version replaced gi:8705171.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             -------------- Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             -------------- Project Information
                 Center project name: L9203
                 Center clone name: 87_E_22
             -------------- Summary Statistics
                 Sequencing vector: M13; M77815; 81% of reads
                 Sequencing vector: Plasmid; n/a; 19% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.960731
                 Consensus quality: 157025 bases at least Q40
                 Consensus quality: 160134 bases at least Q30
                 Consensus quality: 161093 bases at least Q20
```

```
              Insert size: 158000; agarose-fp
              Insert size: 161795; sum-of-contigs
              Quality coverage: 5.4 in Q20 bases; agarose-fp
              Quality coverage: 5.2 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    15243: contig of 15243 bp in length
            *   15244 15343: gap of      100 bp
            *   15344    18897: contig of 3554 bp in length
            *   18898 18997: gap of      100 bp
            *   18998    25211: contig of 6214 bp in length
            *   25212 25311: gap of      100 bp
            *   25312    34838: contig of 9527 bp in length
            *   34839 34938: gap of      100 bp
            *   34939    45351: contig of 10413 bp in length
            *   45352 45451: gap of      100 bp
            *   45452    57259: contig of 11808 bp in length
            *   57260 57359: gap of      100 bp
            *   57360    67990: contig of 10631 bp in length
            *   67991 68090: gap of      100 bp
            *   68091    83399: contig of 15309 bp in length
            *   83400 83499: gap of      100 bp
            *   83500   105395: contig of 21896 bp in length
            *  105396 105495: gap of      100 bp
            *  105496   129959: contig of 24464 bp in length
            *  129960 130059: gap of      100 bp
            *  130060   162795: contig of 32736 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162795
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-87E22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .15243
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15344. .18897
                     /note="assembly_fragment"
     misc_feature    18998. .25211
                     /note="assembly_fragment"
     misc_feature    25312. .34838
                     /note="assembly_fragment"
     misc_feature    34939. .45351
                     /note="assembly_fragment"
     misc_feature    45452. .57259
                     /note="assembly_fragment"
     misc_feature    57360. .67990
                     /note="assembly_fragment"
     misc_feature    68091. .83399
                     /note="assembly_fragment"
     misc_feature    83500. .105395
                     /note="assembly_fragment"
     misc_feature    105496. .129959
                     /note="assembly_fragment"
     misc_feature    130060. .162795
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    42087 a  39347 c  38304 g  42057 t   1000 others
ORIGIN


alignment_scores:
           Quality:  109.00            Length:     22
             Ratio:  4.955              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-215-435-225_COPY_1_22 x AC037441/rev   ..

 Align seg 1/1  to reverse of: AC037441  from: 1  to: 162795

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  159302 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 159253

      17 tMetValValThrGly 22
         ||||||||||||||||
  159252 GATGGTGGTCACTGGA 159237

seq_name: gb_htg17:AC084081

seq_documentation_block:
LOCUS       AC084081   169645 bp    DNA             HTG       11-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-89M8 map 8, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC084081
VERSION     AC084081.3  GI:13273380
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-89M8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2001 this sequence version replaced gi:12583824.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11307
                Center clone name: 89_M_8
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162624 bases at least Q40
                Consensus quality: 165742 bases at least Q30
```

```
            Consensus quality: 167046 bases at least Q20
            Insert size: 169000; agarose-fp
            Insert size: 168045; sum-of-contigs
            Quality coverage: 5.9 in Q20 bases; agarose-fp
            Quality coverage: 5.9 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 17 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1      335: contig of 335 bp in length
          *     336 435: gap of      100 bp
          *     436     2289: contig of 1854 bp in length
          *    2290 2389: gap of      100 bp
          *    2390     4997: contig of 2608 bp in length
          *    4998 5097: gap of      100 bp
          *    5098    64677: contig of 59580 bp in length
          *   64678 64777: gap of      100 bp
          *   64778    68151: contig of 3374 bp in length
          *   68152 68251: gap of      100 bp
          *   68252    72655: contig of 4404 bp in length
          *   72656 72755: gap of      100 bp
          *   72756    78893: contig of 6138 bp in length
          *   78894 78993: gap of      100 bp
          *   78994    84664: contig of 5671 bp in length
          *   84665 84764: gap of      100 bp
          *   84765    90853: contig of 6089 bp in length
          *   90854 90953: gap of      100 bp
          *   90954    98425: contig of 7472 bp in length
          *   98426 98525: gap of      100 bp
          *   98526   106371: contig of 7846 bp in length
          *  106372 106471: gap of      100 bp
          *  106472   115673: contig of 9202 bp in length
          *  115674 115773: gap of      100 bp
          *  115774   127138: contig of 11365 bp in length
          *  127139 127238: gap of      100 bp
          *  127239   138436: contig of 11198 bp in length
          *  138437 138536: gap of      100 bp
          *  138537   149766: contig of 11230 bp in length
          *  149767 149866: gap of      100 bp
          *  149867   167340: contig of 17474 bp in length
          *  167341 167440: gap of      100 bp
          *  167441   169645: contig of 2205 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-89M8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .335
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    436. .2289
                     /note="assembly_fragment"
     misc_feature    2390. .4997
                     /note="assembly_fragment"
     misc_feature    5098. .64677
                     /note="assembly_fragment"
     misc_feature    64778. .68151
                     /note="assembly_fragment"
     misc_feature    68252. .72655
                     /note="assembly_fragment"
     misc_feature    72756. .78893
                     /note="assembly_fragment"
     misc_feature    78994. .84664
                     /note="assembly_fragment"
     misc_feature    84765. .90853
                     /note="assembly_fragment"
     misc_feature    90954. .98425
                     /note="assembly_fragment"
     misc_feature    98526. .106371
                     /note="assembly_fragment"
     misc_feature    106472. .115673
                     /note="assembly_fragment"
     misc_feature    115774. .127138
                     /note="assembly_fragment"
     misc_feature    127239. .138436
                     /note="assembly_fragment"
     misc_feature    138537. .149766
                     /note="assembly_fragment"
     misc_feature    149867. .167340
                     /note="assembly_fragment"
     misc_feature    167441. .169645
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44253 a  41074 c  39720 g  42960 t   1638 others
ORIGIN


alignment_scores:
             Quality:  109.00              Length:     22
               Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC084081/rev   ..

 Align seg 1/1  to reverse of: AC084081  from: 1  to: 169645

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   49659 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 49610

      17 tMetValValThrGly 22
         ||||||||||||||||
   49609 GATGGTGGTCACTGGA 49594

seq_name: gb_pr2:AC005723

seq_documentation_block:
LOCUS       AC005723   127766 bp    DNA             PRI       28-SEP-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 194j15 (LBNL H137) - BAC clone
            35G8 (LBNL H138) - BAC clone 194o13 (LBNL H106),  complete
            sequence.
ACCESSION   AC005723 AC001452 AC001453 AC001454 AC001455 AC002172
VERSION     AC005723.1  GI:3660463
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127766)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127766)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 127766)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1998) Human Genome Center, DOE Joint Genome
```

```
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .127766
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5p"
                     /clone="194o13"
                     /chromosome="5"
                     /note="LBNL H106"
     repeat_region   708. .731
                     /note="(ATTT)6"
                     /rpt_type=tandem
                     /rpt_unit=ATTT
     misc_feature    complement(1725. .1847)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   3041. .3141
                     /rpt_family="Alu"
     repeat_region   3124. .3172
                     /note="(A)49"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   5834. .5857
                     /note="(CCAT)6"
                     /rpt_type=tandem
                     /rpt_unit=CCAT
     repeat_region   complement(6739. .6979)
                     /rpt_family="L1"
     repeat_region   complement(7817. .8218)
                     /rpt_family="THE1"
     repeat_region   10688. .10771
                     /rpt_family="MIR"
     misc_feature    11403. .11542
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   12080. .12159
                     /rpt_family="MIR"
     repeat_region   complement(15524. .16379)
                     /rpt_family="MER4"
     repeat_region   16447. .16574
                     /rpt_family="MSTc"
     repeat_region   16638. .16710
                     /rpt_family="MLTa"
     repeat_region   16755. .18167
                     /rpt_family="MSTAR"
     repeat_region   18558. .18711
                     /rpt_family="MER5"
     repeat_region   complement(21242. .21667)
                     /rpt_family="MLT1"
     repeat_region   complement(21670. .22014)
                     /rpt_family="THE1"
     repeat_region   complement(22807. .22937)
                     /rpt_family="MER42"
     repeat_region   24739. .24778
                     /note="(AT)20"
                     /rpt_type=tandem
                     /rpt_unit=AT
     repeat_region   24779. .24806
                     /note="(AC)14"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   complement(25120. .25532)
                     /rpt_family="LOR1"
     repeat_region   complement(26165. .26444)
                     /rpt_family="Alu"
     repeat_region   26883. .27169
                     /rpt_family="Alu"
     repeat_region   27254. .27474
                     /rpt_family="Alu"
     misc_feature    complement(27668. .27708)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(29011. .29154)
                     /rpt_family="Alu"
     repeat_region   complement(29463. .29767)
                     /rpt_family="Alu"
     repeat_region   complement(30117. .30354)
                     /rpt_family="MER42"
     repeat_region   complement(31144. .31475)
                     /rpt_family="MER2"
     misc_feature    complement(31781. .31847)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(31991. .32229)
                     /rpt_family="L1"
     repeat_region   34182. .34480
                     /rpt_family="Alu"
     repeat_region   36177. .36461
                     /rpt_family="Alu"
     repeat_region   complement(37934. .38017)
                     /rpt_family="Alu"
     repeat_region   38200. .38560
                     /rpt_family="THE1"
     repeat_region   39245. .39468
                     /rpt_family="Alu"
     repeat_region   complement(39823. .40063)
                     /rpt_family="THE1"
     repeat_region   complement(40073. .40420)
                     /rpt_family="MSTAR"
     repeat_region   complement(40425. .40807)
                     /rpt_family="THE1"
     misc_feature    42856. .42977
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(43093. .43800)
                     /rpt_family="L1"
     repeat_region   43803. .45685
                     /rpt_family="L1"
     repeat_region   45747. .47771
                     /standard_name="endogenous retrovirus"
                     /rpt_family="HERVH"
     repeat_region   47837. .48990
                     /standard_name="endogenous retrovirus"
                     /rpt_family="HERVH"
     repeat_region   49149. .49403
                     /standard_name="endogenous retrovirus"
                     /rpt_family="HERVH"
     repeat_region   49518. .49975
                     /standard_name="endogenous retrovirus"
                     /rpt_family="HERVH"
     repeat_region   49976. .50407
                     /rpt_family="LTR7"
     repeat_region   52589. .52941
                     /rpt_family="MIR2"
     repeat_region   complement(52657. .52793)
                     /rpt_family="Alu"
     repeat_region   complement(53015. .53080)
                     /rpt_family="L1"
     repeat_region   54921. .55022
                     /rpt_family="L1"
     repeat_region   56827. .56847
                     /note="(T)21"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(56829. .57122)
                     /rpt_family="Alu"
     repeat_region   57181. .57288
                     /rpt_family="L1"
     repeat_region   58088. .58379
                     /rpt_family="Alu"
     repeat_region   complement(59600. .59892)
                     /rpt_family="Alu"
     repeat_region   60336. .60367
                     /note="(GT)16"
                     /rpt_type=tandem
                     /rpt_unit=GT
     repeat_region   complement(60803. .60881)
                     /rpt_family="MIR"
```

```
     repeat_region   complement(60883. .61891)
                     /rpt_family="LTR8"
     repeat_region   complement(62470. .63175)
                     /rpt_family="MER44C"
     repeat_region   complement(64427. .64612)
                     /rpt_family="MIR"
     repeat_region   65110. .65263
                     /rpt_family="MER58B"
     repeat_region   65358. .65425
                     /rpt_family="Cheshire"
     repeat_region   65422. .65491
                     /rpt_family="MER58"
     repeat_region   complement(66797. .67059)
                     /rpt_family="Alu"
     repeat_region   67827. .67995
                     /rpt_family="MIR"
     repeat_region   complement(69113. .69282)
                     /rpt_family="MLT1"
     repeat_region   complement(71015. .71187)
                     /rpt_family="MST"
     repeat_region   71296. .71397
                     /rpt_family="MSTc"
     repeat_region   72769. .75031
                     /rpt_family="L1"
     repeat_region   75043. .75347
                     /rpt_family="Alu"

alignment_scores:
            Quality:   58.00              Length:     21
              Ratio:  3.222                Gaps:      0
 Percent Similarity:  85.714    Percent Identity: 47.619

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC005723   ..

 Align seg 1/1  to: AC005723  from: 1  to: 127766

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||   |||   :::|||   ||||||:::::::::::::::::::||
   76832 GGGTGGCCGATGATGATAGTAATGGCAGCTGTGGTGATGGCGGTGATAAT 76881

      18 tValValThrGly 22
         |||||||:::|||
   76882 GGTAGTGGTGGGG 76894

seq_name: gb_htg1:AC008437

seq_documentation_block:
LOCUS       AC008437   169700 bp    DNA             HTG       26-SEP-2000
DEFINITION  Homo sapiens chromosome 5 clone CTC-329A22, WORKING DRAFT SEQUENCE,
            17 unordered pieces.
ACCESSION   AC008437
VERSION     AC008437.5  GI:10305077
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169700)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169700)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:8575875.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 311537, H389
            Center clone name: CIT-HSPC_329A22
            -------
            Summary Statistics
            Consensus quality: 152054 bases at least Q40
            Consensus quality: 160606 bases at least Q30
            Consensus quality: 162232 bases at least Q20
            Estimated insert size: 156000; pulse field gel estimation
            Estimated insert size: 168100; sum-of-contigs estimation
            Quality coverage:  6.27 in Q20 bases; pulse field gel estimation
            Quality coverage:  5.82 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1042: contig of 1042 bp in length
            *     1043     1142: gap of unknown length
            *     1143     2192: contig of 1050 bp in length
            *     2193     2292: gap of unknown length
            *     2293     3380: contig of 1088 bp in length
            *     3381     3480: gap of unknown length
            *     3481     4986: contig of 1506 bp in length
            *     4987     5086: gap of unknown length
            *     5087     6137: contig of 1051 bp in length
            *     6138     6237: gap of unknown length
            *     6238     9208: contig of 2971 bp in length
            *     9209     9308: gap of unknown length
            *     9309    13759: contig of 4451 bp in length
            *    13760    13859: gap of unknown length
            *    13860    17412: contig of 3553 bp in length
            *    17413    17512: gap of unknown length
            *    17513    22391: contig of 4879 bp in length
            *    22392    22491: gap of unknown length
            *    22492    25367: contig of 2876 bp in length
            *    25368    25467: gap of unknown length
            *    25468    28671: contig of 3204 bp in length
            *    28672    28771: gap of unknown length
            *    28772    35739: contig of 6968 bp in length
            *    35740    35839: gap of unknown length
            *    35840    44103: contig of 8264 bp in length
            *    44104    44203: gap of unknown length
            *    44204    52953: contig of 8750 bp in length
            *    52954    53053: gap of unknown length
            *    53054    78366: contig of 25313 bp in length
            *    78367    78466: gap of unknown length
            *    78467   112256: contig of 33790 bp in length
            *   112257   112356: gap of unknown length
            *   112357   169700: contig of 57344 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169700
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-329A22"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    52716 a  30825 c  30465 g  52092 t   3602 others
ORIGIN


alignment_scores:
            Quality:   58.00              Length:     21
              Ratio:  3.222                Gaps:      0
 Percent Similarity:  85.714    Percent Identity: 47.619

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC008437/rev   ..

 Align seg 1/1  to reverse of: AC008437  from: 1  to: 169700
```

```
      2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
        ||||||   |||   :::|||   ||||||:::::::::::::::::::||
  158916 GGGTGGCCGATGATGATAGTAATGGCAGCTGTGGTGATGGCGGTGATAAT 158867

     18 tValValThrGly 22
        |||||||:::|||
  158866 GGTAGTGGTGGGG 158854

seq_name: gb_htg1:AC008588

seq_documentation_block:
LOCUS       AC008588   206211 bp    DNA             HTG       05-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone CTC-573M9, WORKING DRAFT SEQUENCE,
            12 unordered pieces.
ACCESSION   AC008588
VERSION     AC008588.4  GI:7708969
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206211)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206211)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 5, 2000 this sequence version replaced gi:7454137.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 405509
            Center clone name: CIT-HSPC_573M9
            -------
            Summary Statistics
            Consensus quality: 189471 bases at least Q40
            Consensus quality: 200343 bases at least Q30
            Consensus quality: 202679 bases at least Q20
            Estimated insert size: 210000; pulse field gel estimation
            Estimated insert size: 205111; sum-of-contigs estimation
            Quality coverage:  5.15 in Q20 bases; pulse field gel estimation
            Quality coverage:  5.28 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2056: contig of 2056 bp in length
            *    2057     2156: gap of unknown length
            *    2157     7014: contig of 4858 bp in length
            *    7015     7114: gap of unknown length
            *    7115    11487: contig of 4373 bp in length
            *   11488    11587: gap of unknown length
            *   11588    16807: contig of 5220 bp in length
            *   16808    16907: gap of unknown length
            *   16908    24738: contig of 7831 bp in length
            *   24739    24838: gap of unknown length
            *   24839    35601: contig of 10763 bp in length
            *   35602    35701: gap of unknown length
            *   35702    51573: contig of 15872 bp in length
            *   51574    51673: gap of unknown length
            *   51674    70235: contig of 18562 bp in length
            *   70236    70335: gap of unknown length
            *   70336    89224: contig of 18889 bp in length
            *   89225    89324: gap of unknown length
            *   89325   114744: contig of 25420 bp in length
            *  114745   114844: gap of unknown length
            *  114845   140424: contig of 25580 bp in length
            *  140425   140524: gap of unknown length
            *  140525   206211: contig of 65687 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206211
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-573M9"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    62241 a  39339 c  37513 g  66015 t   1103 others
ORIGIN


alignment_scores:
            Quality:   58.00            Length:     21
              Ratio:  3.222              Gaps:      0
 Percent Similarity:  85.714   Percent Identity: 47.619

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC008588/rev   ..

 Align seg 1/1  to reverse of: AC008588  from: 1  to: 206211

      2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
        ||||||   |||   :::|||   ||||||:::::::::::::::::::||
  134156 GGGTGGCCGATGATGATAGTAATGGCAGCTGTGGTGATGGCGGTGATAAT 134107

     18 tValValThrGly 22
        |||||||:::|||
  134106 GGTAGTGGTGGGG 134094

seq_name: gb_ro1:MUSNECDIN

seq_documentation_block:
LOCUS       MUSNECDIN    1585 bp    mRNA            ROD       27-APR-1993
DEFINITION  M.musculus necdin mRNA, complete cds.
ACCESSION   M80840
VERSION     M80840.1  GI:200019
KEYWORDS    brain-specific protein; neural differentiation; nuclear protein.
SOURCE      Mus musculus (strain C3H/He) (library: lambda gt10) embryo cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1585)
  AUTHORS   Maruyama,K., Usami,M., Aizawa,T. and Yoshikawa,K.
  TITLE     A novel brain-specific mRNA encoding nuclear protein (necdin)
            expressed in neurally differentiated embryonal carcinoma cells
  JOURNAL   Biochem. Biophys. Res. Commun. 178, 291-296 (1991)
  MEDLINE   91298962
FEATURES             Location/Qualifiers
     source          1. .1585
                     /organism="Mus musculus"
                     /strain="C3H/He"
                     /db_xref="taxon:10090"
                     /cell_line="P19(P19S1801A1)"
                     /cell_type="carcinoma"
                     /dev_stage="embryo"
                     /tissue_lib="lambda gt10"
     CDS             68. .1045
                     /codon_start=1
                     /evidence=experimental
                     /product="necdin"
                     /protein_id="AAA39805.1"
                     /db_xref="GI:200020"
                     /translation="MSEQSKDLSDPNFAAEVPDCEMQDSDAVPVGIPPPASLAANLAG
                     PPCAPEGPMAAQQASPPPEERIEDVDPKILQQAAEEGRAHQPQSPARPIPAPPAPAQL
                     VQKAHELMWYVLVKDQKRMVLWFPDMVKEVMGSYKKWCRSILRRTSVILARVFGLHLR
```

```
                    LTNLHTMEFALVKALSPEELDRVALNNRMPMTGLLLMILSLIYVKGRGAREGAVWNVL
                    RILGLRPWKKHSTFGDVRKIITEEFVQQNYLKYQRVPHIEPPEYEFFWGSRANREITK
                    MQIMEFLARVFKKDPQAWPSRYREALEQARALREANLAAQAPRSSVSED"
     polyA_signal   1564. .1569
     polyA_site     1585
BASE COUNT      380 a    422 c    436 g    347 t
ORIGIN


alignment_scores:
            Quality:   55.00             Length:     16
              Ratio:  4.231               Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 68.750

alignment_block:
 US-09-215-435-225_COPY_1_22 x MUSNECDIN/rev   ..

 Align seg 1/1  to reverse of: MUSNECDIN  from: 1  to: 1585

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMet 17
         ||||||||||:::   ||||||   ||||||||||||   ||||||:::
     337 GGCTGGACTCTGGGGCTGGTGGGCGCGGCCCTCCTCTGCGGCCTGCTG 290

seq_name: gb_ro1:MUSNCP

seq_documentation_block:
LOCUS        MUSNCP       3899 bp    DNA             ROD       04-FEB-1999
DEFINITION   Mouse gene for necdin, complete cds and promoter sequence.
ACCESSION    D76440
VERSION      D76440.1  GI:1040690
KEYWORDS     necdin; nuclear protein.
SOURCE       Mus musculus (strain:129SV) DNA, clone_lib:library of Stratagene
             clone:lambda-nd5.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 3899)
  AUTHORS    Yoshikawa,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-OCT-1995) to the DDBJ/EMBL/GenBank databases. Kazuaki
             Yoshikawa, Institute for Protein Research, Osaka University, Div.
             Regulation of Macromolecular Functions; Yamadaoka 3-2, Suita, Osaka
             565, Japan (E-mail:yoshikaw@protein.osaka-u.ac.jp, Tel:06-879-8621,
             Fax:06-879-8623)
REFERENCE    2  (bases 1 to 3899)
  AUTHORS    Maruyama,K., Usami,M., Aizawa,T. and Yoshikawa,K.
  TITLE      A novel brain-specific mRNA encoding nuclear protein (necdin)
             expressed in neurally differentiated embryonal carcinoma cells
  JOURNAL    Biochem. Biophys. Res. Commun. 178 (1), 291-296 (1991)
  MEDLINE    91298962
REFERENCE    3  (bases 1 to 2520)
  AUTHORS    Uetsuki,T., Takagi,K., Sugiura,H. and Yoshikawa,K.
  TITLE      Structure and expression of the mouse necdin gene. Identification
             of a postmitotic neuron-restrictive core promoter
  JOURNAL    J. Biol. Chem. 271 (2), 918-924 (1996)
  MEDLINE    96132826
REFERENCE    4  (sites)
  AUTHORS    Kuo,C.H., Uetsuki,T., Kim,C.H., Tanaka,H., Li,B.S., Taira,E.,
             Higuchi,H., Okamoto,H., Yoshikawa,K. and Miki,N.
  TITLE      Determination of a necdin cis-acting element required for neuron
             specific expression by using zebra fish
  JOURNAL    Biochem. Biophys. Res. Commun. 211 (2), 438-446 (1995)
  MEDLINE    95314610
FEATURES             Location/Qualifiers
     source          1. .3899
                     /organism="Mus musculus"
                     /strain="129SV"
                     /db_xref="taxon:10090"
                     /clone="lambda-nd5"
                     /clone_lib="library of Stratagene"
     promoter        791. .836
                     /standard_name="necdin core promoter"
                     /note="The core promoter contains a consensus motif of
                     xenobiotic responsive elements (XRE) and a GC-rich
                     region."
                     /citation=[3]
                     /citation=[4]
                     /function="postmitotic neuron-specific expression"
                     /evidence=experimental
     exon            871. .2478
     mRNA            871. .2478
     5'UTR           871. .956
     CDS             957. .1934
                     /codon_start=1
                     /product="necdin"
                     /protein_id="BAA11183.1"
                     /db_xref="GI:1040691"
                     /translation="MSEQSKDLSDPNFAAEAPDCEMQDSDAVPVGIPPPASLAANLAG
                     PPCAPEGPMAAQQASPPPEERIEDVDPKILQQAAEEGRAHQPQSPARPIPAPPAPAQL
                     VQKAHELMWYVLVKDQKRMVLWFPDMVKEVMGSYKKWCRSILRRTSVILARVFGLHLR
                     LTNLHTMEFALVKALSPEELDRVALNNRMPMTGLLLMILSLIYVKGRGAREGAVWNVL
                     RILGLRPWKKHSTFGDVRKIITEEFVQQNYLKYQRVPHIEPPEYEFFWGSRANREITK
                     MQIMEFLARVFKKDPQAWPSRYREALEQARALREANLAAQAPRSSVSED"
     variation       1006
                     /note="This variation gives rise to amino acid
                     substitution (AA17: Ala to Val)"
                     /citation=[2]
                     /replace="t"
     3'UTR           1935. .2478
     polyA_site      2478
BASE COUNT     1107 a    895 c    892 g   1005 t
ORIGIN


alignment_scores:
            Quality:   55.00             Length:     16
              Ratio:  4.231               Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 68.750

alignment_block:
 US-09-215-435-225_COPY_1_22 x MUSNCP/rev   ..

 Align seg 1/1  to reverse of: MUSNCP  from: 1  to: 3899

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMet 17
         ||||||||||:::   ||||||   ||||||||||||   ||||||:::
    1226 GGCTGGACTCTGGGGCTGGTGGGCGCGGCCCTCCTCTGCGGCCTGCTG 1179

seq_name: gb_htg11:AC026388_0

seq_documentation_block:
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC026388 Accession AC026388
     Fragment Name      Begin     End
     AC026388_0             1     110000
     AC026388_1        100001     210000
     AC026388_2        200001     310000
     AC026388_3        300001     410000
     AC026388_4        400001     431552
LOCUS        AC026388   431552 bp    DNA             HTG       07-JAN-2001
DEFINITION   Mus musculus chromosome 7 clone RP23-426B15, WORKING DRAFT
             SEQUENCE, 28 unordered pieces.
ACCESSION    AC026388
VERSION      AC026388.16  GI:12025600
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 431552)
  AUTHORS    Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
             Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
             Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
             Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
             Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
             Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
```

```
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 431552)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 4, 2001 this sequence version replaced gi:11096426.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MADM
                Center clone name: RP23-426B15
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 75% of reads
                Chemistry: Dye-terminator Big Dye: 25% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 410961 bases at least Q40
                Consensus quality: 420079 bases at least Q30
                Consensus quality: 424939 bases at least Q20
                Estimated insert size: 413976; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    86039: contig of 86039 bp in length
            *    86040    86139: gap of unknown length
            *    86140   143869: contig of 57730 bp in length
            *   143870   143969: gap of unknown length
            *   143970   177956: contig of 33987 bp in length
            *   177957   178056: gap of unknown length
            *   178057   225623: contig of 47567 bp in length
            *   225624   225723: gap of unknown length
            *   225724   260078: contig of 34355 bp in length
            *   260079   260178: gap of unknown length
            *   260179   294645: contig of 34467 bp in length
            *   294646   294745: gap of unknown length
            *   294746   323628: contig of 28883 bp in length
            *   323629   323728: gap of unknown length
            *   323729   349798: contig of 26070 bp in length
            *   349799   349898: gap of unknown length
            *   349899   368486: contig of 18588 bp in length
            *   368487   368586: gap of unknown length
            *   368587   381021: contig of 12435 bp in length
            *   381022   381121: gap of unknown length
            *   381122   388190: contig of 7069 bp in length
            *   388191   388290: gap of unknown length
            *   388291   394160: contig of 5870 bp in length
            *   394161   394260: gap of unknown length
            *   394261   400308: contig of 6048 bp in length
            *   400309   400408: gap of unknown length
            *   400409   405084: contig of 4676 bp in length
            *   405085   405184: gap of unknown length
            *   405185   408985: contig of 3801 bp in length
            *   408986   409085: gap of unknown length
            *   409086   413664: contig of 4579 bp in length
            *   413665   413764: gap of unknown length
            *   413765   416758: contig of 2994 bp in length
            *   416759   416858: gap of unknown length
            *   416859   417971: contig of 1113 bp in length
            *   417972   418071: gap of unknown length
            *   418072   419956: contig of 1885 bp in length
            *   419957   420056: gap of unknown length
            *   420057   421336: contig of 1280 bp in length
            *   421337   421436: gap of unknown length
            *   421437   422711: contig of 1275 bp in length
            *   422712   422811: gap of unknown length
            *   422812   423845: contig of 1034 bp in length
            *   423846   423945: gap of unknown length
            *   423946   425267: contig of 1322 bp in length
            *   425268   425367: gap of unknown length
            *   425368   426834: contig of 1467 bp in length
            *   426835   426934: gap of unknown length
            *   426935   428003: contig of 1069 bp in length
            *   428004   428103: gap of unknown length
            *   428104   429170: contig of 1067 bp in length
            *   429171   429270: gap of unknown length
            *   429271   430288: contig of 1018 bp in length
            *   430289   430388: gap of unknown length
            *   430389   431552: contig of 1164 bp in length.
FEATURES             Location/Qualifiers
     source          1. .431552
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="RP23-426B15"
BASE COUNT   125430 a  89222 c  89661 g 124505 t   2734 others
ORIGIN


alignment_scores:
            Quality:   55.00              Length:     16
              Ratio:  4.231                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 68.750

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC026388_0/rev   ..

 Align seg 1/1  to reverse of: AC026388_0  from: 1  to: 110000

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMet 17
         ||||||||||:::   ||||||   ||||||||||||   ||||||:::
   74226 GGCTGGACTCTGGGGCTGGTGGGCGCGGCCCTCCTCTGCGGCCTGCTG 74179

seq_name: gb_ro1:AC027298

seq_documentation_block:
LOCUS       AC027298   201467 bp    DNA             ROD       05-DEC-2000
DEFINITION  Mus musculus 7 BAC RP23-266F22 (Roswell Park Cancer Institute Mouse
            BAC Library) complete sequence.
ACCESSION   AC027298
VERSION     AC027298.20  GI:11496331
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201467)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
```

Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R., Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M., Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S., Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R., Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S., Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B., Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S., Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A., Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G., Worley,K. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201467)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 201467)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 201467)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 1, 2000 this sequence version replaced gi:11321700. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES Location/Qualifiers

source	1. .201467 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="7" /clone="RP23-266F22"
repeat_region	complement(54. .289) /rpt_family="MULV"
repeat_region	895. .1313 /rpt_family="MTC"
repeat_region	1762. .1785 /rpt_family="AT_rich"
repeat_region	2815. .2852 /rpt_family="AT_rich"
repeat_region	complement(3115. .4840) /rpt_family="Lx2"
repeat_region	4836. .4870 /rpt_family="L1"
repeat_region	complement(4912. .5804) /rpt_family="Lx5"
repeat_region	5879. .6078 /rpt_family="Lx4"
repeat_region	complement(6132. .7418) /rpt_family="Lx4"
repeat_region	7429. .7451 /rpt_family="AT_rich"
repeat_region	7654. .8483 /rpt_family="L1_MM"
repeat_region	8568. .8648 /rpt_family="L1"
repeat_region	complement(19374. .19511) /rpt_family="Lx"
repeat_region	19516. .19994 /rpt_family="L1_MM"
repeat_region	complement(20009. .20523) /rpt_family="L1F"
repeat_region	complement(20527. .20825) /rpt_family="L1VL2"
repeat_region	complement(20835. .21490) /rpt_family="Lx4"
repeat_region	complement(21486. .22307) /rpt_family="Lx4"
repeat_region	22308. .22337 /rpt_family="(TTTG)n"
repeat_region	complement(24534. .24625) /rpt_family="L1"
repeat_region	complement(24623. .24743) /rpt_family="L1"
repeat_region	24745. .26002 /rpt_family="L1_MM"
repeat_region	26091. .26259 /rpt_family="RLTR10"
repeat_region	29240. .29697 /rpt_family="L1_MM"
repeat_region	complement(30087. .30773) /rpt_family="ETnERV2"
repeat_region	complement(31847. .32382) /rpt_family="MMETn"
repeat_region	complement(32387. .33027) /rpt_family="ETnERV2"
repeat_region	complement(33013. .33334) /rpt_family="ETnERV2"
repeat_region	complement(33349. .33402) /rpt_family="ETnERV2"
repeat_region	33403. .33884 /rpt_family="LTRIS_MM"
repeat_region	complement(33885. .33949) /rpt_family="ETnERV2"
repeat_region	complement(33997. .34265) /rpt_family="ETnERV2"
misc_feature	complement(34577. .35032) /note="Region similar to Mm#S176249 vb56h02.r1 Mus musculus cDNA: gb=AA387402"
repeat_region	complement(35348. .35455) /rpt_family="ETnERV2"
repeat_region	35456. .35642 /rpt_family="L1_MM"
repeat_region	36292. .36343 /rpt_family="(CATATA)n"
repeat_region	36378. .36460 /rpt_family="A-rich"
repeat_region	complement(36461. .36757) /rpt_family="Lx7"
repeat_region	complement(36850. .38900) /rpt_family="Lx2"
repeat_region	complement(38914. .40129) /rpt_family="Lx2"
repeat_region	complement(40132. .40249) /rpt_family="Lx2"
repeat_region	40592. .41250 /rpt_family="L1"
repeat_region	41396. .41541 /rpt_family="L1"
repeat_region	41777. .41917 /rpt_family="L1"
repeat_region	complement(41944. .42321) /rpt_family="Lx2"
repeat_region	complement(42481. .43067) /rpt_family="Lx2"

```
     repeat_region   49133. .50140
                     /rpt_family="RLTR13D"
     repeat_region   54195. .54661
                     /rpt_family="L1_MM"
     repeat_region   complement(56789. .58214)
                     /rpt_family="L1VL2"
     repeat_region   58666. .58786
                     /rpt_family="HERVL"
     repeat_region   59692. .60261
                     /rpt_family="MT-INTERNAL"
     repeat_region   60314. .60343
                     /rpt_family="(TTTA)n"
     repeat_region   complement(60438. .62610)
                     /rpt_family="Lx2"
     repeat_region   65129. .65223
                     /rpt_family="RMER20"
     repeat_region   65340. .65610
                     /rpt_family="Lx5"
     repeat_region   complement(65954. .66079)
                     /rpt_family="RMER20"
     repeat_region   66080. .66418
                     /rpt_family="Lx5"
     repeat_region   66419. .66458
                     /rpt_family="(CA)n"
     repeat_region   66459. .66836
                     /rpt_family="Lx5"
     repeat_region   66899. .67131
                     /rpt_family="Lx7"
     repeat_region   67148. .67201
                     /rpt_family="(TG)n"
     repeat_region   complement(67264. .67400)
                     /rpt_family="Lx"
     repeat_region   complement(67402. .67684)
                     /rpt_family="ORR1A2"
     repeat_region   67794. .67965
                     /rpt_family="B2"
     repeat_region   68188. .68440
                     /rpt_family="Lx9"
     misc_feature    complement(68713. .69166)

alignment_scores:
            Quality:   55.00            Length:     16
              Ratio:  4.231              Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 68.750

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC027298   ..

 Align seg 1/1  to: AC027298  from: 1  to: 201467

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMet 17
         ||||||||||:::   ||||||   ||||||||||||   ||||||:::
  196746 GGCTGGACTCTGGGGCTGGTGGGCGCGGCCCTCCTCTGCGGCCTGCTG 196793

seq_name: gb_ov:ORZHSC70

seq_documentation_block:
LOCUS       ORZHSC70     2261 bp    mRNA            VRT       13-APR-1999
DEFINITION  Oryzias latipes mRNA for heat shock protein 70 cognate, complete
            cds.
ACCESSION   D13669
VERSION     D13669.1  GI:538526
KEYWORDS    heat shock protein 70 cognate.
SOURCE      Oryzias latipes (strain:HB32) fibroblast-like cell_line:cultured
            Medaka cell OL32 cDNA to mRNA.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 2261)
  AUTHORS   Arai,A., Naruse,K., Mitani,H. and Shima,A.
  TITLE     Cloning and characterization of cDNAs for 70-kDa heat-shock
            proteins (Hsp70) from two fish species of the genus Oryzias
  JOURNAL   Jpn. J. Genet. 70 (3), 423-433 (1995)
  MEDLINE   95358880
REFERENCE   2  (bases 1 to 2261)
  AUTHORS   Naruse,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1992) to the DDBJ/EMBL/GenBank databases. Kiyoshi
            Naruse, School of Science, University of Tokyo., Department of
            Biological Sciences; Hongo 7-3-1, Bunkyo-Ku, Tokyo 113, Japan
            (E-mail:naruse@biol.s.u-tokyo.ac.jp, Tel:03-3812-2111(ex.4443),
            Fax:03-3816-1965)
FEATURES             Location/Qualifiers
     source          1. .2261
                     /organism="Oryzias latipes"
                     /strain="HB32"
                     /db_xref="taxon:8090"
                     /cell_line="cultured Medaka cell OL32"
                     /cell_type="fibroblast-like"
     CDS             122. .2182
                     /codon_start=1
                     /product="heat shock protein 70 cognate"
                     /protein_id="BAA76887.1"
                     /db_xref="GI:4589737"
                     /translation="MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAF
                     TDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDDHVVQSDMNDWPFNVINDNTRPK
                     VQVEYKGETKSFYPEEVSSMVLTKMKEIAEAYLGKTVNNAVITVPAYFNDSQRQATKD
                     AGTISGLNVLRIINEPTAAAIAYGLDKKVGSERNVLIFDLGGGTFDVSILTIEDGIFE
                     VKSTAGDTHLGGEDFDNRMVNHFIAEFKRKYKKDISDNKRAVRRLRSACERAKRTLSS
                     STQASIEIDSLYEGVDFYTSITRARFEELNADLFRGTLDPVEKSLRDAKMDKGQIHDI
                     VLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENVQDLL
                     LLDVTPLSLGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLNQVYEVGAMTK
                     DNNLLGKFELTGIPPALWCPQIEVTFDIDANGIMNVSAVDKSTGKENKITITNDKGRL
                     SKEDIERMVQEAEKYKAEDDVQRDKVSAKNGLESYAFNMKSTVEDEKLAGKISDEDKQ
                     KILDKCNEVISWLDKNQTAERDEYVHQQKELEKVCNPIITKLYQSAGGMPGGCQRECQ
                     EVSLELVVLLAVALLDQPLKKLIKHSMTSPSKIFTKKTLFKCNMQSIKSVMRLNQFPF
                     ITIT"
BASE COUNT      647 a    496 c    538 g    580 t
ORIGIN


alignment_scores:
            Quality:   54.00            Length:     20
              Ratio:  3.176              Gaps:      0
 Percent Similarity:  85.000   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x ORZHSC70   ..

 Align seg 1/1  to: ORZHSC70  from: 1  to: 2261

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||   ||||||:::      :::::::::|||:::|||::
     669 ATGGGCTGGACAAAAAGGTTGGCTCCGAGAGGAATGTTCTTATCTTTGAT 718

      17 tMetValVal 20
         ::::||||||
     719 CTTGGTGGTG 728

seq_name: gb_pl2:AF193060

seq_documentation_block:
LOCUS       AF193060     3937 bp    DNA             PLN       12-NOV-2000
DEFINITION  Phanerochaete chrysosporium NADPH-dependent cytochrome P450
            oxidoreductase (CPR) gene, complete cds.
ACCESSION   AF193060
VERSION     AF193060.1  GI:11138368
KEYWORDS    .
SOURCE      Phanerochaete chrysosporium.
  ORGANISM  Phanerochaete chrysosporium
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Hymenomycetidae;
            Aphyllophorales; Corticiaceae; Phanerochaete.
REFERENCE   1  (bases 1 to 3937)
  AUTHORS   Yadav,J.S. and Loper,J.C.
```

```
  TITLE     Cytochrome P450 oxidoreductase gene and its differentially
            terminated cDNAs from the white rot fungus Phanerochaete
            chrysosporium
  JOURNAL   Curr. Genet. 37 (1), 65-73 (2000)
  MEDLINE   20136378
   PUBMED   10672447
REFERENCE   2  (bases 1 to 3937)
  AUTHORS   Yadav,J.S. and Loper,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1999) Environmental Health, University of
            Cincinnati, 3223 Eden Avenue, Cincinnati, OH 45267-0056, USA
FEATURES             Location/Qualifiers
     source          1. .3937
                     /organism="Phanerochaete chrysosporium"
                     /strain="BKMF-1767"
                     /db_xref="taxon:5306"
     mRNA            join(<992. .1177,1240. .1807,1858. .2834,2893. .3572)
                     /gene="CPR"
                     /note="alternatively polyadenylated"
                     /product="NADPH-dependent cytochrome P450 oxidoreductase"
     mRNA            join(<992. .1177,1240. .1807,1858. .2834,2893. .3402)
                     /gene="CPR"
                     /note="alternatively polyadenylated"
                     /product="NADPH-dependent cytochrome P450 oxidoreductase"
     gene            <992. .>3572
                     /gene="CPR"
     CDS             join(992. .1177,1240. .1807,1858. .2834,2893. .3372)
                     /gene="CPR"
                     /EC_number="1.6.2.4"
                     /codon_start=1
                     /product="NADPH-dependent cytochrome P450 oxidoreductase"
                     /protein_id="AAG31349.1"
                     /db_xref="GI:11138369"
                     /translation="MAVSSSSDVIVLSVGIILAALYLFREQIFSAAKPKTVQVPSSKA
                     AAGGNGNPRDFIAKMKEGKKRIVIFYGSQTGTAEEYAIRLAKEAKSKFGLASLVCDPE
                     EYDFENLDQVPEDCCVFFVMATYGEGEPTDNAVQLCQNLSDESFEFSNGEHKLPGLKY
                     VIFGLGNKTYEHYNLISRNVDRDLQKMGAIRIGERGEGDDDKSMEEDYLEWKDGMWEA
                     FAKAMNVEEGQGGDSPDFVVTEVFDHPEEKVYLGELSARALTRTKGIHDAKNPYPAPI
                     IAAKELFAPGSDRNCVHIELSTESSGITYQHGDHVGVWPSNADKEVDRLLYALGLHEK
                     KDTVINIESLDPALAKVPFPVPTTYATVLRHYIDISALAGRQILGVLAKFAPNPEAEA
                     VLKDLNSNKEHYQNIVANGCMKLGEVLQYAAGNDLHADPTASNTTAWKIPFDIIVSSI
                     PRLQPRYYSISSSPKLYPNAIHATVVVLKYKSEKAPRVEERWIYGVGSNFLLNLKYAS
                     HHDKAATLVSDDSPSEPSIVSHYPTYSIEGPRGAYKQGDVVKVPIHVRRSTFRLPTNP
                     KSPVIMIGPGTGVAPFRGFVQERVAMARRTIEKHGPEGLADWGPIRLYYGCRRSDQDF
                     LYKDEWPEYAKELHGKFIMRCAFSREPPYKPDGSKIYVQDLIWEDAEQIADAILNGKG
                     YVYICGDAKSMSKSVEETLCRILGEAKGGSAEVEGAAELKLLKERNRLLLDVWS"
BASE COUNT      875 a   1092 c   1073 g    897 t
ORIGIN


alignment_scores:
            Quality:   54.00             Length:     16
              Ratio:  4.154               Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 68.750

alignment_block:
 US-09-215-435-225_COPY_1_22 x AF193060   ..

 Align seg 1/1  to: AF193060  from: 1  to: 3937

      2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMet 17
        ||||||||||:::|||      |||||||||   ||||||||||||:::
    707 GGCTGGACTTTGCGAAACGAAACCGCTGCCTGGCTCCTCGGTTTAGTT 754

seq_name: gb_htg4:AC014412

seq_documentation_block:
LOCUS       AC014412    86446 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014412
VERSION     AC014412.1  GI:6436923
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 86446)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212722 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .86446
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    23099 a  20123 c  20064 g  23160 t
ORIGIN


alignment_scores:
            Quality:   54.00             Length:     17
              Ratio:  3.375               Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 58.824

alignment_block:
 US-09-215-435-225_COPY_1_22 x AC014412   ..

 Align seg 1/1  to: AC014412  from: 1  to: 86446

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ::::::|||:::|||||||||||||||:::::   :::|||||||||||||||
  14131 TTGAGTTGGGTAATGAGACTAGTTGTATGCAATATTTTGCTTGGGTTAAT 14180

     17 t 17
        |
  14181 G 14181
```

THIS PAGE BLANK (USPTO)

OM of: US-09-215-435-225_COPY_1_22 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:29 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582	+	109.00	303.87	5.9e-09	447	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480	+	109.00	299.97	9.7e-09	681	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660	+	109.00	298.18	1.2e-08	826	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013	+	109.00	298.18	1.2e-08	826	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252	+	109.00	298.18	1.2e-08	826	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350	+	109.00	297.49	1.3e-08	890	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490	+	109.00	297.42	1.4e-08	897	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582	+	109.00	297.36	1.4e-08	903	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539	+	109.00	297.30	1.4e-08	909	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882	+	109.00	297.06	1.4e-08	932	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431	+	109.00	296.87	1.5e-08	952	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426	+	109.00	296.16	1.6e-08	1028	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13175	-	49.00	93.40	3.1e+03	29729	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV44217	-	48.00	124.31	59.29	777	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV52281	+	48.00	98.85	1.6e+03	12127	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA31710	+	47.00	126.29	46.00	461	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34773	+	47.00	125.78	49.10	487	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC45669	+	47.00	125.35	51.87	510	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC32882	+	47.00	125.32	52.11	512	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT28519	-	47.00	120.68	94.53	845	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT42063	-	47.00	49.50	6.6e+05	183012	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ34737	+	46.50	91.46	4.0e+03	16941	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78455	+	46.00	122.77	72.24	495	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78544	+	46.00	121.76	82.24	552	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF71597	+	46.00	104.49	753.89	3561	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ39542	+	46.00	103.93	809.31	3780	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAV74399	+	45.50	102.25	1.0e+03	3886	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX98102	-	45.00	121.77	82.13	405	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC52772	+	45.00	119.56	109.03	514	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC95279	+	45.00	118.95	117.90	549	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX98101	+	45.00	115.88	174.90	765	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ52569	+	45.00	115.02	195.18	839	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF55661	+	45.00	113.46	238.46	993	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX85781	+	45.00	107.70	499.23	1849	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX98035	-	45.00	78.71	2.0e+04	42235	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC99043	+	44.50	108.87	429.79	1397	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ88155	+	44.50	104.95	709.92	2131	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV89602	+	44.00	126.63	44.02	176	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV89600	+	44.00	125.31	52.15	203	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC74745	+	44.00	120.53	96.27	340	!
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA48965	+	44.00	117.13	149.00	491	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT12701	+	44.00	114.19	217.12	674	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ80075	+	44.00	110.52	347.83	100	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT93249	+	44.00	107.42	517.61	140	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAD01349	+	44.00	106.96	548.96	147	

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582

seq_documentation_block:
ID AAX40582 standard; cDNA; 447 BP.
XX
AC AAX40582;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID No: 182.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906550-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153780/13.
DR P-PSDB; AAY11860.
XX
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 1; Page 285; 675pp; English.
XX
CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 447 BP; 98 A; 110 C; 136 G; 99 T; 4 other;

alignment_scores:
Quality: 109.00 Length: 22
Ratio: 4.955 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
 US-09-215-435-225_COPY_1_22 x AAX40582   ..

 Align seg 1/1  to: AAX40582  from: 1  to: 447

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 134

      17 tMetValValThrGly 22
         ||||||||||||||||
     135 GATGGTGGTCACTGGA 150

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480

seq_documentation_block:
ID   AAF94480 standard; cDNA; 681 BP.
XX
AC   AAF94480;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Human hydrophobic domain containing protein clone HP03880 cDNA #104.
XX
KW   Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW   antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW   cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW   HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW   inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW   behavioural characteristic; immune response; ss.
XX
OS   Homo sapiens.
XX
PN   WO200112660-A2.
XX
PD   22-FEB-2001.
XX
PF   10-AUG-2000; 2000WO-JP05356.
XX
PR   17-AUG-1999;   99JP-0230344.
PR   07-SEP-1999;   99JP-0252551.
PR   01-OCT-1999;   99JP-0281132.
PR   22-OCT-1999;   99JP-0301624.
PR   04-NOV-1999;   99JP-0313877.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-160059/16.
DR   P-PSDB; AAB88590.
XX
PT   Human proteins with hydrophobic domains and the DNAs which encode them
PT   are useful for treating autoimmune disorders, burns and tumors and for
PT   screening novel pharmaceuticals -
XX
PS   Claim 3; Page 426-427; 518pp; English.
XX
CC   AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC   AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC   anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC   osteopathic, anti-inflammatory and cytostatic activities, and can be
CC   used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC   to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC   probes for genetic diagnosis and gene sources for gene therapy or for
CC   producing (I) in large quantities. Cells containing (II) are used for
CC   the detection of ligands or receptors corresponding to membrane or
CC   secretory proteins and to screen small molecule novel pharmaceuticals.
CC   Antibodies directed to (I) can be used for the detection, quantification
CC   and purification of (I). Activities of (I) may include cytokine and cell
CC   proliferation/differentiation function, immune stimulating or suppressing
CC   activity, haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC   activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC   multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC   inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC   wound healing, as nutritional sources or supplements e.g. as amino acid,
CC   carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 681 BP; 179 A; 186 C; 183 G; 133 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAF94480   ..

 Align seg 1/1  to: AAF94480  from: 1  to: 681

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 50

      17 tMetValValThrGly 22
         ||||||||||||||||
      51 GATGGTGGTCACTGGA 66

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660

seq_documentation_block:
ID   AAX97660 standard; DNA; 826 BP.
XX
AC   AAX97660;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 124.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY35976.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 255; 516pp; English.
```

```
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAX97660   ..

 Align seg 1/1  to: AAX97660  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGly 22
         ||||||||||||||||
      65 GATGGTGGTCACTGGA 80

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013

seq_documentation_block:
ID   AAC00013 standard; cDNA; 826 BP.
XX
AC   AAC00013;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein cDNA sequence #4.
XX
KW   Human; secreted protein; 5' EST; expressed sequence tag; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             15..698
FT                   /*tag= a
FT                   /product= "secreted protein"
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00016.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Example 19; SEQ ID 7; 71pp + CD-ROM; English.
XX
CC   The present sequence is a full length cDNA encoding a human
CC   secreted protein. The cDNA was obtained from a 5' EST using first and
CC   second strand synthesis procedures. 5' ESTs were prepared from total
CC   human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region
CC   (UTR) of the mRNA because they are often obtained from oligo-dT primed
CC   cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC   sequences derived from the 5' ends of mRNAs and even in those cases where
CC   longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC   included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC   therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC   are also used in diagnostic, forensic, gene therapy and chromosome
CC   mapping procedures. They are used to obtain upstream regulatory sequences
CC   and to design expression and secretion vectors.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAC00013   ..

 Align seg 1/1  to: AAC00013  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGly 22
         ||||||||||||||||
      65 GATGGTGGTCACTGGA 80

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252

seq_documentation_block:
ID   AAZ42252 standard; cDNA; 826 BP.
XX
AC   AAZ42252;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human phosphatidylethanolamine-binding protein encoding cDNA.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
```

```
DR   P-PSDB; AAY64647.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Example 21; Page 168-169; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAZ42252   ..

 Align seg 1/1  to: AAZ42252  from: 1  to: 826

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      15 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 64

      17 tMetValValThrGly 22
         ||||||||||||||||
      65 GATGGTGGTCACTGGA 80

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350

seq_documentation_block:
ID   AAA96350 standard; cDNA; 890 BP.
XX
AC   AAA96350;
XX
DT   08-FEB-2001  (first entry)
XX
DE   cDNA encoding a novel polypeptide designated PRO4408.
XX
KW   Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW   PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW   PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW   PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW   insulinemia; kidney disorder; Bergers disease; nephropathy;
KW   Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW   Crohns disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             89..760
FT                   /*tag=  a
FT   sig_peptide     89..154
FT                   /*tag=  b
XX
PN   WO200056889-A2.
XX
PD   28-SEP-2000.
XX
PF   01-MAR-2000; 2000WO-US05601.
XX
PR   23-MAR-1999;   99US-0125774.
PR   23-MAR-1999;   99US-0125778.
PR   24-MAR-1999;   99US-0125826.
PR   31-MAR-1999;   99US-0127035.
PR   05-APR-1999;   99US-0127706.
PR   21-APR-1999;   99US-0130359.
PR   27-APR-1999;   99US-0131270.
PR   27-APR-1999;   99US-0131272.
PR   27-APR-1999;   99US-0131291.
PR   04-MAY-1999;   99US-0132371.
PR   04-MAY-1999;   99US-0132379.
PR   04-MAY-1999;   99US-0132383.
PR   25-MAY-1999;   99US-0135750.
PR   08-JUN-1999;   99US-0138166.
PR   20-JUL-1999;   99US-0144791.
PR   03-AUG-1999;   99US-0146970.
PR   09-DEC-1999;   99US-0170262.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2000-628263/60.
DR   P-PSDB; AAB18923.
XX
PT   Novel secreted and transmembrane polypeptides useful for diagnosing
PT   tumour in a mammal, for identifying agonists and antagonists of the
PT   polypeptide and for therapeutic use
XX
PS   Claim 2; Fig 29; 222pp; English.
XX
CC   The present sequence encodes a secreted or transmembrane polypeptide.
CC   The specification describes polypeptides designated PRO1484, PRO4334,
CC   PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC   PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC   PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC   useful for diagnosing tumour in a mammal. The polypeptides, their
CC   agonists and antagonists are useful treating a condition associated with
CC   expression or activity of the polypeptide. Conditions treated include
CC   obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC   capable of inducing proliferation of mammalian kidney mesangial cells
CC   and are therefore useful for treating kidney disorders associated with
CC   decreased mesangial cell function such as Bergers disease or other
CC   nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC   dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC   to generate transgenic animals for use in development and screening of
CC   therapeutically useful reagents and also for chromosome identification
CC   and tissue typing.
XX
SQ   Sequence 890 BP; 228 A; 246 C; 234 G; 182 T; 0 other;


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAA96350   ..

 Align seg 1/1  to: AAA96350  from: 1  to: 890

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       89 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 138

       17 tMetValValThrGly 22
          ||||||||||||||||
      139 GATGGTGGTCACTGGA 154
```

```
seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490

seq_documentation_block:
ID   AAF94490 standard; cDNA; 897 BP.
XX
AC   AAF94490;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Human hydrophobic domain containing protein clone HP03880 cDNA #114.
XX
KW   Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW   antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW   cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW   HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW   inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW   behavioural characteristic; immune response; ss.
XX
OS   Homo sapiens.
XX
PN   WO200112660-A2.
XX
PD   22-FEB-2001.
XX
PF   10-AUG-2000; 2000WO-JP05356.
XX
PR   17-AUG-1999;   99JP-0230344.
PR   07-SEP-1999;   99JP-0252551.
PR   01-OCT-1999;   99JP-0281132.
PR   22-OCT-1999;   99JP-0301624.
PR   04-NOV-1999;   99JP-0313877.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-160059/16.
DR   P-PSDB; AAB88590.
XX
PT   Human proteins with hydrophobic domains and the DNAs which encode them
PT   are useful for treating autoimmune disorders, burns and tumors and for
PT   screening novel pharmaceuticals -
XX
PS   Claim 4; Page 442-444; 518pp; English.
XX
CC   AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC   AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC   anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC   osteopathic, anti-inflammatory and cytostatic activities, and can be
CC   used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC   to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC   probes for genetic diagnosis and gene sources for gene therapy or for
CC   producing (I) in large quantities. Cells containing (II) are used for
CC   the detection of ligands or receptors corresponding to membrane or
CC   secretory proteins and to screen small molecule novel pharmaceuticals.
CC   Antibodies directed to (I) can be used for the detection, quantification
CC   and purification of (I). Activities of (I) may include cytokine and cell
CC   proliferation/differentiation function, immune stimulating or suppressing
CC   activity, haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC   activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC   multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC   inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC   wound healing, as nutritional sources or supplements e.g. as amino acid,
CC   carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 897 BP; 226 A; 253 C; 234 G; 184 T; 0 other;


alignment_scores:
           Quality:  109.00              Length:     22
             Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAF94490   ..

 Align seg 1/1  to: AAF94490  from: 1  to: 897

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 148

      17 tMetValValThrGly 22
         ||||||||||||||||
     149 GATGGTGGTCACTGGA 164

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582

seq_documentation_block:
ID   AAA15582 standard; cDNA; 903 BP.
XX
AC   AAA15582;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human phospholipid binding protein 2, PLBP2 gene.
XX
KW   Human; phospholipid binding protein; PLBP2; foetal development disorder;
KW   reproduction disorder; cell proliferation disorder; immune response;
KW   autoimmune disorder; AIDS; infertility; cytostatic; immunosuppressive;
KW   gene therapy; hereditary neuropathy;
KW   phosphatidylethanolamine binding protein D1; PE-BP D1; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..771
FT                   /*tag=  a
FT                   /product= "Human PLBP2"
XX
PN   US6063767-A.
XX
PD   16-MAY-2000.
XX
PF   09-DEC-1998;   98US-0208718.
XX
PR   28-OCT-1997;   97US-0958820.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Corley NC,  Shah P,  Lal P,  Hillman JL;
XX
DR   WPI; 2000-375529/32.
DR   P-PSDB; AAY94263.
XX
PT   New purified phospholipid binding proteins 1 and 2 useful for
PT   diagnosing, treating or preventing diseases disorders associated with
PT   fetal development, reproduction, cell proliferation, and the immune
PT   response  -
XX
PS   Example 5; Fig 2; 37pp; English.
XX
```

```
CC   The present sequence is the phospholipid binding protein 2 (PLBP2) gene.
CC   This gene is expressed in lung, prostate and heart tissues. Also, the
CC   protein is expressed in foetal tumour tissues. PLBP2 may be used for the
CC   diagnosis, prevention, or treatment of disorders associated with foetal
CC   development (e.g. hereditary neuropathies), reproduction (e.g.
CC   infertility), cell proliferation (e.g. cancers), and the immune response
CC   (AIDS). PLBP2 antibodies may also be developed for potential drug
CC   screening or to quantitate PLBP2 gene expression in biopsied tissues. The
CC   PLBP2 gene may be administered for gene therapy of disorders associated
CC   with PLBP2. PLBP2 has high homology with the phosphatidylethanolamine
CC   binding protein D1, PE-BP D1, of Onchocerca volvulus. PE-BP DI is thought
CC   to play a role in transport or signal mechanisms between membranes and
CC   the cytoplasm.
XX
SQ   Sequence 903 BP; 222 A; 251 C; 242 G; 188 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAA15582   ..

 Align seg 1/1  to: AAA15582  from: 1  to: 903

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 137

      17 tMetValValThrGly 22
         ||||||||||||||||
     138 GATGGTGGTCACTGGA 153

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539

seq_documentation_block:
ID   AAC77539 standard; cDNA; 909 BP.
XX
AC   AAC77539;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF3094 polynucleotide sequence SEQ ID NO:6187.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43330.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5369-5370; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 909 BP; 232 A; 254 C; 236 G; 186 T; 1 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAC77539   ..

 Align seg 1/1  to: AAC77539  from: 1  to: 909

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     100 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 149

      17 tMetValValThrGly 22
         ||||||||||||||||
     150 GATGGTGGTCACTGGA 165

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882

seq_documentation_block:
ID   AAC84882 standard; cDNA; 932 BP.
XX
AC   AAC84882;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human SEC1 nucleic acid sequence (clone ID 3445452).
XX
KW   SECX; cytostatic; gynecological; gene therapy; screening assay; human;
KW   chromosomal mapping; forensic biology; cell proliferation; cancer;
KW   cell differentiation; immune associated disorder; gestational disease;
KW   SEC1; ss.
XX
```

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             113..781
FT                   /*tag= a
FT                   /product= "SEC1"
FT   sig_peptide     113..181
FT                   /*tag= b
FT   mat_peptide     182..778
FT                   /*tag= c
XX
PN   WO200078802-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17328.
XX
PR   23-JUN-1999;   99US-0140584.
PR   20-JUL-1999;   99US-0144722.
PR   16-SEP-1999;   99US-0154520.
PR   22-JUN-2000; 2000US-0604286.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Vernet C,  Yang M,  Boldog FL;
PI   Herrmann JL;
XX
DR   WPI; 2001-071385/08.
DR   P-PSDB; AAB48368.
XX
PT   Polynucleotides encoding SECX proteins useful for treating disease
PT   characterized by an aberrant level of cell proliferation and/or
PT   differentiation like cancer or immune associated disorders -
XX
PS   Claim 3; Fig 1; 132pp; English.
XX
CC   The invention relates to human SECX polypeptides and polynucleotides
CC   encoding them. The SECX polypeptides can be expressed by standard
CC   recombinant methodology.  The SECX polypeptides are useful for treating
CC   or preventing a SECX-associated disorder. The invention is useful in
CC   screening assays; detection assays (e.g. chromosomal mapping, cell and
CC   tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC   prognostic assays, monitoring clinical trials, and pharmocogenomics); and
CC   methods of treatment (e.g. therapeutic and prophylactic), especially
CC   disorders characterized by aberrant cell proliferation and/ or
CC   differentiation like cancer or immune associated disorders or gestational
CC   disease. The present sequence represents a SEC1 nucleic acid sequence.
XX
SQ   Sequence 932 BP; 244 A; 257 C; 244 G; 187 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAC84882   ..

 Align seg 1/1  to: AAC84882  from: 1  to: 932

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     113 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 162

      17 tMetValValThrGly 22
         |||||||||||||||
     163 GATGGTGGTCACTGGA 178

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431

seq_documentation_block:
ID   AAA78431 standard; cDNA; 952 BP.
XX
AC   AAA78431;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:61.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24487.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 478; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 952 BP; 310 A; 247 C; 223 G; 172 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAA78431   ..

 Align seg 1/1  to: AAA78431  from: 1  to: 952

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     63 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 112

     17 tMetValValThrGly 22
        ||||||||||||||||
    113 GATGGTGGTCACTGGA 128

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426

seq_documentation_block:
ID   AAA78426 standard; cDNA; 1028 BP.
XX
AC   AAA78426;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:56.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24482.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 475; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1028 BP; 309 A; 268 C; 259 G; 186 T; 6 other;


alignment_scores:
          Quality:  109.00            Length:     22
             Ratio:  4.955                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAA78426    ..

 Align seg 1/1  to: AAA78426  from: 1  to: 1028

      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    153 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 202

     17 tMetValValThrGly 22
        ||||||||||||||||
    203 GATGGTGGTCACTGGA 218

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13175

seq_documentation_block:
ID   AAX13175 standard; DNA; 29729 BP.
XX
AC   AAX13175;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:238.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1181-1196; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 29729 BP; 9748 A; 4927 C; 6290 G; 8752 T; 12 other;
```

```
alignment_scores:
            Quality:   49.00              Length:     18
              Ratio:  3.267                Gaps:      0
 Percent Similarity:  83.333   Percent Identity: 55.556

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAX13175/rev   ..

 Align seg 1/1  to reverse of: AAX13175  from: 1  to: 29729

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         :::   |||:::   :::||||||::::::|||||||||||||   |||||
   10490 TTAATGCTTTTACCAGGAGCGTTAGTCATGGGACTTATGATGCCAGTAAC 10441

      21 rGly 22
         ||||
   10440 TGGT 10437

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV44217

seq_documentation_block:
ID   AAV44217 standard; DNA; 777 BP.
XX
AC   AAV44217;
XX
DT   19-JUL-1999  (first entry)
XX
DE   Lettuce resistance gene AC15-2B.
XX
KW   Resistance gene; AC15-2B; RG gene; lettuce; disease resistance;
KW   pest resistance; virus; fungus; protozoon; bacterium; nematode;
KW   crop protection; transgenic plant; fingerprinting; ss.
XX
OS   Lactuca sativa.
XX
PN   WO9830083-A1.
XX
PD   16-JUL-1998.
XX
PF   09-JAN-1998;   98WO-US00615.
XX
PR   10-JAN-1997;   97US-0781734.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Meyers B,  Michelmore RW,  Shen K;
XX
DR   WPI; 1998-398692/34.
XX
PT   New resistance gene nucleic acids - useful to produce disease
PT   resistant plants, e.g. lettuce, Lactuca sativa, and to detect
PT   resistance genes, e.g. to fingerprint cultivars
XX
PS   Disclosure; Page 93-97; 183pp; English.
XX
CC   This is the nucleotide sequence of the lettuce AC15-2B gene
CC   which hybridises to lettuce RG2 resistance genes (see AAV44199-214)
CC   of the invention.  The invention relates to families of resistance
CC   genes, termed RG or RLG, from lettuce.  RG families RG1-RG5 and RG7
CC   have been identified, each RG family being defined as a group of
CC   polypeptide sequences that have at least 60% amino acid sequence
CC   identity.  Individual members of an RG family typically map to the
CC   same genomic locus.  RG nucleic acid sequences (see AAV44188-257) can
CC   be used to confer resistance in plants to a variety of pests,
CC   including viruses, fungi, nematodes, insects and bacteria.
CC   Sequences from within the RG genes can be used to fingerprint
CC   cultivars or germplasm for the presence of desired resistance
CC   genes.  Promoters of RG genes can be used to drive heterologous
CC   gene expression under conditions in which RG genes are expressed.
CC   RG polynucleotides can also be used to produce antisense and
CC   ribozyme molecules useful for inhibiting RG activity in plants.
CC   The invention also provides RG proteins (see AAW64150-98) that are
CC   encoded by these RG genes.
XX
SQ   Sequence 777 BP; 261 A; 148 C; 120 G; 246 T; 2 other;


alignment_scores:
            Quality:   48.00              Length:     19
              Ratio:  3.200                Gaps:      0
 Percent Similarity:  78.947   Percent Identity: 42.105

alignment_block:
 US-09-215-435-225_COPY_1_22 x AAV44217/rev   ..

 Align seg 1/1  to reverse of: AAV44217  from: 1  to: 777

       3 TrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetVa 19
         |||   :::|||:::||||||   :::||||||:::||||||   :::
     577 TGGTTATTGCGAGTTGTTACCAATTCTCTACTTGTTGGACTCTCACTCTC 528

      19 lValThr 21
         ::::::
     527 TATCTCA 521

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV52281

seq_documentation_block:
ID   AAV52281 standard; DNA; 12127 BP.
XX
AC   AAV52281;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:148.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 991-998; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
```

CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 12127 BP; 3743 A; 2512 C; 2345 G; 3527 T; 0 other;

alignment_scores:
Quality: 48.00 Length: 21
Ratio: 2.824 Gaps: 0
Percent Similarity: 80.952 Percent Identity: 47.619

alignment_block:
US-09-215-435-225_COPY_1_22 x AAV52281 ..

Align seg 1/1 to: AAV52281 from: 1 to: 12127

```
       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         :::|||      |||   ::::::||||||:::|||||||||||||:::||
    1479 TTAGGAAATGGCATGAGTGTCCTAACCGCTTCGTTACTATTAGGAATAAT 1528

      17 tMetValValThr 21
         |::::::   |||
    1529 GATTTTGCCAACC 1541
```

```
OM of: US-09-215-435-225_COPY_1_22 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  5:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4 +  109.00  318.52  4.3e-10   903  !  S
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-05320A-1342 +   47.00  135.16  7.05      3
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-743-637B-25 -   47.00  127.65  18.46     845  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-526-840B-25 -   47.00  127.65  18.46     845  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-198-122-1 +   44.50  111.08  154.71    2124  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-328-111-159 +   44.00  116.75  74.73     1002  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-307-896-2 -   44.00  112.77  124.47    1516  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-344-914-1 -   44.00  112.77  124.47    1516  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-11808-2 -   44.00  112.51  128.73    1558
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-054-077C-1 -   44.00  104.97  338.57    3415  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-226-568-18 +   44.00  103.61  403.05    3934  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-077-848A-1 +   44.00  103.58  404.57    3946  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-211-640-1 +   44.00  103.58  404.57    3946  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-378-536-1 +   44.00  103.58  404.57    3946  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-03547-1 +   44.00  103.58  404.57    3946
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-471-119A-1 -   44.00   79.80  8.5e+03   46899  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2 +   44.00   36.16  6.7e+05   4403765
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-1 +   44.00   36.15  6.7e+05   4411529
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-166-346A-4 -   43.00  125.64  23.90      288  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-171-982-1 -   43.00  123.34  32.11      366  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-091-405-1 -   43.00  112.11  135.56    1178  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-00909-3 +   43.00  111.60  144.69    1242
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-148-215A-3 -   43.00  109.67  185.27    1518  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-580-680-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-480-156A-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-354-961-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-581-094-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-580-665-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-581-142-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-581-103-2 -   43.00  106.33  284.50    2150  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-05442A-2 -   43.00  106.33  284.50    2150
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-944-449-3 -   43.00  106.32  284.66    2151  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-09-027-337-9 +   43.00  103.45  411.35    2900  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-09-027-337-1 +   43.00  102.67  454.94    3147  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-812-7 +   43.00   98.17  809.61    5024  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-827-7 +   43.00   98.17  809.61    5024  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-921-177-7 +   43.00   98.17  809.61    5024  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-362-577C-7 +   43.00   98.17  809.61    5024  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-920-828-7 +   43.00   98.17  809.61    5024  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-149-103-1 -   42.50  100.58  594.60    3330  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-451-883-1 -   42.50  100.58  594.60    3330
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-393-734-1 -   42.50   99.68  667.12    3656
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-894-489-1 -   42.50   99.68  667.12    3656
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-393-734-3 -   42.50   90.42  2.2e+03   9592
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-894-489-3 -   42.50   90.42  2.2e+03   9592

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4

seq_documentation_block:
; Sequence 4, Application US/09208718
; Patent No. 6063767
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer
;    APPLICANT:  Corley, Neil
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN PHOSPHOLIPID BINDING PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,718
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/958,820
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0379 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  LUNGTUT12
;      CLONE:  3126479
US-09-208-718-4


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-208-718-4  ...

 Align seg 1/1  to: US-09-208-718-4  from: 1  to: 903

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 137

      17 tMetValValThrGly 22
         ||||||||||||||||
     138 GATGGTGGTCACTGGA 153
```

```
seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-05320A-1342

seq_documentation_block:
; Sequence 1342, Application PC/TUS9605320A
;  GENERAL INFORMATION:
;    APPLICANT:  Human Genome Sciences
;    APPLICANT:  9410 Key West Avenue
;    APPLICANT:  Rockville, MD 20850
;    APPLICANT:  United States of America
;    APPLICANT:  Johns Hopkins University
;    APPLICANT:  720 Rutland Avenue
;    APPLICANT:  Baltimore, MD 21205
;    APPLICANT:  United States of America
;    APPLICANT:  Mark D. Adams
;    APPLICANT:  Owen White
;    APPLICANT:  Hamilton O. Smith
;    APPLICANT:  J. Craig Venter
;    TITLE OF INVENTION:  Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20003-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/05320A
;      FILING DATE:  April22, 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/476,102
;      FILING DATE:  June 7, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,429
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eric K. Steffe
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.014PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  1342:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US96-05320A-1342


alignment_scores:
            Quality:   47.00              Length:     18
              Ratio:  3.357                Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 55.556

alignment_block:
 US-09-215-435-225_COPY_1_22 x PCT-US96-05320A-1342   ..

 Align seg 1/1  to: PCT-US96-05320A-1342  from: 1  to: 387

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         |||:::   :::|||   ||||||:::||||||      :::|||||||||
      13 ATGAAAAAATTAACATTAGCATTGGTTTTAGGTTCAGCTTTAGTCGTGAC 62

      21 rGly 22
```

```
         ||||
      63 AGGT 66

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-743-637B-25

seq_documentation_block:
; Sequence 25, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  845 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Haemophilus influenzae
US-08-743-637B-25


alignment_scores:
            Quality:   47.00              Length:     18
              Ratio:  3.357                Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 55.556

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-743-637B-25/rev   ..

 Align seg 1/1  to reverse of: US-08-743-637B-25  from: 1  to: 845

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         |||:::   :::|||   ||||||:::||||||      :::|||||||||
     423 ATGAAAAAATTAACATTAGCATTGGTTTTAGGTTCAGCTTTAGTCGTGAC 374

      21 rGly 22
         ||||
     373 AGGT 370
```

```
seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-526-840B-25

seq_documentation_block:
; Sequence 25, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  845 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Haemophilus influenzae
US-08-526-840B-25


alignment_scores:
            Quality:   47.00              Length:     18
              Ratio:  3.357                Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 55.556

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-526-840B-25/rev    ..

 Align seg 1/1  to reverse of: US-08-526-840B-25  from: 1  to: 845

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         |||:::   :::|||   ||||||:::||||||      :::||||||||
     423 ATGAAAAAATTAACATTAGCATTGGTTTTAGGTTCAGCTTTAGTCGTGAC 374

      21 rGly 22
         ||||
     373 AGGT 370

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-198-122-1
```

```
seq_documentation_block:
; Sequence 1, Application US/09198122
; Patent No. 6180380
;  GENERAL INFORMATION:
;    APPLICANT:  Strebhardt, Klaus; Rubsamen-Waigmann, Helga;
;    APPLICANT:  Holtrich, Uwe
;    TITLE OF INVENTION:  CLONING OF A MEMBER OF THE SERINE-
;    TITLE OF INVENTION:  THREONINE-KINASE FAMILY
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate SX-20
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/198,122
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/601,014
;      FILING DATE:  23-FEB-1996
;      APPLICATION NUMBER:  PCT/EP94/02863
;      FILING DATE:  30-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 4329177
;      FILING DATE:  30-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9516-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  yes
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  65..1873
US-09-198-122-1


alignment_scores:
            Quality:   44.50              Length:     20
              Ratio:  3.423                Gaps:      1
 Percent Similarity:  65.000   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-198-122-1    ..

 Align seg 1/1  to: US-09-198-122-1  from: 1  to: 2124

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||||||:::|||      ||||||          |||||||::
    1305 GGGTGGACTATTCGGACAAGTACGGCC.........TTGGGTATCAGCTC 1345

      18 tValValThr 21
```

```
          |||:::|||
    1346 TGTGATAACA 1355

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-328-111-159

seq_documentation_block:
; Sequence 159, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 159
;   LENGTH: 1002
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1002)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-111-159


alignment_scores:
            Quality:   44.00              Length:      17
              Ratio:  3.385                Gaps:       0
 Percent Similarity:  76.471   Percent Identity: 52.941

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-328-111-159    ..

 Align seg 1/1  to: US-09-328-111-159  from: 1  to: 1002

       3 TrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetVa 19
         |||:::      ||||||||||||   |||||||||   :::::::::||
     548 TGGAGTTGCTTTTTGGTAACAGCTCCATTGCTACTCCCCATTTTATTGGT 597

      19 l 19
         |
     598 T 598

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-307-896-2

seq_documentation_block:
; Sequence 2, Application US/08307896C
; Patent No. 6034071
; GENERAL INFORMATION:
;  APPLICANT: Iyengar, Srinivas Ravi
;  TITLE OF INVENTION: MUTANT ACTIVATED GSALPHA AND ADENYLYL
;  TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;  FILE REFERENCE: 29770
;  CURRENT APPLICATION NUMBER: US/08/307,896C
;  CURRENT FILING DATE: 1994-09-16
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1516
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-307-896-2


alignment_scores:
            Quality:   44.00              Length:      16
              Ratio:  3.143                Gaps:       0
 Percent Similarity:  87.500   Percent Identity: 56.250

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-307-896-2/rev    ..

 Align seg 1/1  to reverse of: US-08-307-896-2  from: 1  to: 1516

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeu 16
         :::|||   |||:::|||||||:::   :::|||||||:::|||||||||
      81 GTTGGCCTCACGCTGCGCCTTCTCCTCGTTGCGCTGGTCCTCGGTCTT 34

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-344-914-1

seq_documentation_block:
; Sequence 1, Application US/09344914
; Patent No. 6110664
; GENERAL INFORMATION:
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF G-ALPHA-S1 EXPRESSION
;  FILE REFERENCE: RTS-0068
;  CURRENT APPLICATION NUMBER: US/09/344,914
;  CURRENT FILING DATE: 1999-06-25
;  NUMBER OF SEQ ID NOS: 87
; SEQ ID NO 1
;   LENGTH: 1516
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13)..(1155)
US-09-344-914-1


alignment_scores:
            Quality:   44.00              Length:      16
              Ratio:  3.143                Gaps:       0
 Percent Similarity:  87.500   Percent Identity: 56.250

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-344-914-1/rev    ..

 Align seg 1/1  to reverse of: US-09-344-914-1  from: 1  to: 1516

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeu 16
         :::|||   |||:::|||||||:::   :::|||||||:::|||||||||
      81 GTTGGCCTCACGCTGCGCCTTCTCCTCGTTGCGCTGGTCCTCGGTCTT 34

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-11808-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9511808
;  GENERAL INFORMATION:
;    APPLICANT:  Iyengar, Srinivas Ravi V.
;    TITLE OF INVENTION:  MUTANT ACTIVATED GsALPHA AND
;    TITLE OF INVENTION:  ADENYLYL
;    TITLE OF INVENTION:  CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue and
;      ADDRESSEE:  Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
```

```
;      COUNTRY:  U.S.
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11808
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/307,896
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S.
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  29970 165/28755
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 408-2500
;      TELEFAX:  (212) 765-2519
;      TELEX:  650 6111063
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US95-11808-2


alignment_scores:
            Quality:   44.00              Length:     16
              Ratio:  3.143                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 56.250

alignment_block:
 US-09-215-435-225_COPY_1_22 x PCT-US95-11808-2/rev    ..

 Align seg 1/1  to reverse of: PCT-US95-11808-2  from: 1  to: 1558

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeu 16
         :::|||   |||:::||||||:::   :::|||||||:::||||||||||
      81 GTTGGCCTCACGCTGCGCCTTCTCCTCGTTGCGCTGGTCCTCGGTCTT 34

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-054-077C-1

seq_documentation_block:
; Sequence 1, Application US/08054077C
; Patent No. 5527679
;  GENERAL INFORMATION:
;    APPLICANT:  HEMLER, MARTIN E.
;    APPLICANT:  RAMASWAMY, HEMAVATHI
;    TITLE OF INVENTION:  HUMAN INTEGRIN  5 SUBUNIT PROTEIN
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 WATER STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/054,077C
;      FILING DATE:  27-APR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/694314
;      FILING DATE:  01-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNICK, DAVID S.
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  40937
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 523-3400
;      TELEFAX:  (617) 523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3415 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  mat peptide
;      LOCATION:  406..2733
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  337..2733
US-08-054-077C-1


alignment_scores:
            Quality:   44.00              Length:     18
              Ratio:  2.750                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 44.444

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-054-077C-1/rev    ..

 Align seg 1/1  to reverse of: US-08-054-077C-1  from: 1  to: 3415

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||   ::::::||||||   ::::::::|||:::||||||:::::
    1290 GGATGGATAGTCCATCTGGTTGGATGCTGTGTACTCGTTGGCCTCGTTCA 1241

      18 tVal 19
         :|||
    1240 GGTG 1237

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-226-568-18

seq_documentation_block:
; Sequence 18, Application US/09226568
; Patent No. 6001992
; GENERAL INFORMATION:
;  APPLICANT: Ackermann, Elizabeth J.
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Marcusson, Eric G.
;  TITLE OF INVENTION: Antisense Modulation of No. 6001992el Anti-apoptotic
;  TITLE OF INVENTION: bcl-2-Related Proteins
;  FILE REFERENCE: ISPH-0337
;  CURRENT APPLICATION NUMBER: US/09/226,568
;  CURRENT FILING DATE: 1999-01-07
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 3934
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: antisense
;   OTHER INFORMATION: sequence
US-09-226-568-18


alignment_scores:
            Quality:   44.00              Length:     23
```

```
            Ratio:  2.444                Gaps:      1
 Percent Similarity:  78.261   Percent Identity: 56.522

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-226-568-18   ..

 Align seg 1/1  to: US-09-226-568-18  from: 1  to: 3934

       1 MetGlyTrpThrMetArgLeuVal..ThrAlaAlaLeuLeuLeuGlyLeu 16
         |||:::|||      |||||||||  |||   :::|||||||||||||||
    2417 ATGGCTTGGAAAAGCAGGCTAGTCTAACCATGGTGCTATTATTAGGCTTG 2466

      17 MetMetValValThrGly 22
         :::::::::   ||||||
    2467 CTTGTTACACACACAGGT 2484

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-077-848A-1

seq_documentation_block:
; Sequence 1, Application US/08077848A
; Patent No. 5470955
;  GENERAL INFORMATION:
;    APPLICANT:  Craig, Ruth W.
;    TITLE OF INVENTION:  ANTIBODIES WHICH SPECIFICALLY BIND mcl-1
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/077,848A
;      FILING DATE:  16-JUN-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  PD-2845
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3946 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  mcl-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  61..1110
;      OTHER INFORMATION:  /note= "When nucleotide 740 = C,
;      OTHER INFORMATION:  amino acid 227 = A; when nucleotide 740 = T, amino
;      OTHER INFORMATION:  acid 227 = V."
 US-08-077-848A-1


 alignment_scores:
             Quality:   44.00              Length:      23
               Ratio:  2.444                Gaps:       1
  Percent Similarity:  78.261   Percent Identity: 56.522

 alignment_block:
```

```
 US-09-215-435-225_COPY_1_22 x US-08-077-848A-1   ..

 Align seg 1/1  to: US-08-077-848A-1  from: 1  to: 3946

       1 MetGlyTrpThrMetArgLeuVal..ThrAlaAlaLeuLeuLeuGlyLeu 16
         |||:::|||      |||||||||  |||   :::|||||||||||||||
    2417 ATGGCTTGGAAAAGCAGGCTAGTCTAACCATGGTGCTATTATTAGGCTTG 2466

      17 MetMetValValThrGly 22
         :::::::::   ||||||
    2467 CTTGTTACACACACAGGT 2484

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-211-640-1

seq_documentation_block:
; Sequence 1, Application US/09211640
; Patent No. 6020466
;  GENERAL INFORMATION:
;    APPLICANT:  Craig, Ruth W.
;    TITLE OF INVENTION:  ANTIBODIES WHICH SPECIFICALLY BIND mcl-1
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/211,640
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/441,375
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  PD-2845
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3946 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  mcl-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  61..1110
;      OTHER INFORMATION:  /note= "When nucleotide 740 = C,
;      OTHER INFORMATION:  amino acid 227 = A; when nucleotide 740 = T, amino
;      OTHER INFORMATION:  acid 227 = V."
US-09-211-640-1


alignment_scores:
            Quality:   44.00              Length:      23
              Ratio:  2.444                Gaps:       1
 Percent Similarity:  78.261   Percent Identity: 56.522

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-211-640-1   ..
```

```
 Align seg 1/1  to: US-09-211-640-1  from: 1  to: 3946

       1 MetGlyTrpThrMetArgLeuVal..ThrAlaAlaLeuLeuLeuGlyLeu 16
         |||:::|||      |||||||||  |||   :::|||||||||||||||
    2417 ATGGCTTGGAAAAGCAGGCTAGTCTAACCATGGTGCTATTATTAGGCTTG 2466

      17 MetMetValValThrGly 22
         :::::::::   ||||||
    2467 CTTGTTACACACACAGGT 2484

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-378-536-1

seq_documentation_block:
; Sequence 1, Application US/09378536
; Patent No. 6200763
;  GENERAL INFORMATION:
;    APPLICANT:  Craig, Ruth W.
;    TITLE OF INVENTION:  ANTIBODIES WHICH SPECIFICALLY BIND mcl-1
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/378,536
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/077,848
;      FILING DATE:  16-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  PD-2845
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3946 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  mcl-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  61..1110
;      OTHER INFORMATION:  /note= "When nucleotide 740 = C,
;      OTHER INFORMATION:  amino acid 227 = A; when nucleotide 740 = T, amino
;      OTHER INFORMATION:  acid 227 = V."
US-09-378-536-1


alignment_scores:
           Quality:   44.00            Length:     23
             Ratio:  2.444              Gaps:      1
 Percent Similarity:  78.261   Percent Identity: 56.522

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-378-536-1   ..
```

```
 Align seg 1/1  to: US-09-378-536-1  from: 1  to: 3946

       1 MetGlyTrpThrMetArgLeuVal..ThrAlaAlaLeuLeuLeuGlyLeu 16
         |||:::|||      |||||||||  |||   :::|||||||||||||||
    2417 ATGGCTTGGAAAAGCAGGCTAGTCTAACCATGGTGCTATTATTAGGCTTG 2466

      17 MetMetValValThrGly 22
         :::::::::   ||||||
    2467 CTTGTTACACACACAGGT 2484

seq_name: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-03547-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9403547
;  GENERAL INFORMATION:
;    APPLICANT:  The Johns Hopkins University School of Medicine
;    TITLE OF INVENTION:  MYELOID CELL LEUKEMIA ASSOCIATED GENE
;    TITLE OF INVENTION:  MCL-1
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03547
;      FILING DATE:  31-MAR-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wetherell, Jr., Ph.D., John W.
;      REGISTRATION NUMBER:  31,678
;      REFERENCE/DOCKET NUMBER:  FD-2845
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3946 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  mcl-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  61..1110
;      OTHER INFORMATION:  /note= "When nucleotide 740 = C,
;      OTHER INFORMATION:  amino acid 227 = A; when nucleotide 740 = T, amino
;      OTHER INFORMATION:  acid 227 = V."
PCT-US94-03547-1


alignment_scores:
           Quality:   44.00            Length:     23
             Ratio:  2.444              Gaps:      1
 Percent Similarity:  78.261   Percent Identity: 56.522

alignment_block:
 US-09-215-435-225_COPY_1_22 x PCT-US94-03547-1   ..

 Align seg 1/1  to: PCT-US94-03547-1  from: 1  to: 3946

       1 MetGlyTrpThrMetArgLeuVal..ThrAlaAlaLeuLeuLeuGlyLeu 16
         |||:::|||      |||||||||  |||   :::|||||||||||||||
    2417 ATGGCTTGGAAAAGCAGGCTAGTCTAACCATGGTGCTATTATTAGGCTTG 2466
```

```
 17 MetMetValValThrGly 22
    :::::::::   ||||||
2467 CTTGTTACACACACAGGT 2484
```

OM of: US-09-215-435-225_COPY_1_22 to: Pending_Patents_NA_Main:* out_format : pfs

Date: Nov 15, 2001 3:55 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-901-904-6667 +			109.00	316.30	2.8e-08	190
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7707 +			109.00	316.30	2.8e-08	19
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7685 +			109.00	314.88	3.4e-08	22
/cgn1_7/ptodata/1/pna/US090_COMB.seq:US-09-035-171-1718 +			109.00	314.26	3.7e-08	239
/cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-212A-10721 +			109.00	314.26	3.7e-08	
/cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-039-416-1718 +			109.00	314.26	3.7e-08	23
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-856-524-163 +			109.00	314.19	3.7e-08	241
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-856-524-905 +			109.00	314.19	3.7e-08	241
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-862-968-1249 +			109.00	314.19	3.7e-08	241
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7694 +			109.00	314.19	3.7e-08	24
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7700 +			109.00	314.19	3.7e-08	24
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7704 +			109.00	314.19	3.7e-08	24
/cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-017-766-163 +			109.00	314.19	3.7e-08	241
/cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-017-766-905 +			109.00	314.19	3.7e-08	241
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-904-3732 +			109.00	314.08	3.8e-08	244
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-904A-3732 +			109.00	314.08	3.8e-08	24
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-957-777-896 +			109.00	314.08	3.8e-08	244
/cgn1_7/ptodata/1/pna/US090_COMB.seq:US-09-035-171-1257 +			109.00	314.08	3.8e-08	244
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7678 +			109.00	314.08	3.8e-08	24
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7681 +			109.00	314.08	3.8e-08	24
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7709 +			109.00	314.08	3.8e-08	24
/cgn1_7/ptodata/1/pna/US6002_COMB.seq:US-60-029-281-896 +			109.00	314.08	3.8e-08	244
/cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-039-416-1257 +			109.00	314.08	3.8e-08	24
/cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-100-454-3236 +			109.00	313.90	3.8e-08	249
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7713 +			109.00	313.90	3.8e-08	24
/cgn1_7/ptodata/1/pna/US6005_COMB.seq:US-60-051-749-3236 +			109.00	313.90	3.8e-08	24
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-904-2031 +			109.00	313.69	4.0e-08	255
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-904A-2031 +			109.00	313.69	4.0e-08	25
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7679 +			109.00	313.69	4.0e-08	25
/cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-013-365-2031 +			109.00	313.69	4.0e-08	25
/cgn1_7/ptodata/1/pna/US091_COMB.seq:US-09-105-427-112 +			109.00	313.42	4.1e-08	263
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-271-490-17026 +			109.00	313.42	4.1e-08	26
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7686 +			109.00	313.42	4.1e-08	26
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-925-552-17026 +			109.00	313.42	4.1e-08	26
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-271-490-17027 +			109.00	313.32	4.1e-08	26
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-925-552-17027 +			109.00	313.32	4.1e-08	26
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-303-031A-3380 +			109.00	312.43	4.6e-08	29
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-894-949-3380 +			109.00	312.43	4.6e-08	294
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7716 +			109.00	312.40	4.7e-08	29
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-385 +			109.00	312.25	4.8e-08	300
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-443A-306 +			109.00	311.85	5.0e-08	
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-903-474-2982 +			109.00	311.85	5.0e-08	
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7701 +			109.00	311.85	5.0e-08	
/cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-012-233-306 +			109.00	311.85	5.0e-08	
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-726-176-478 +			109.00	311.82	5.0e-08	

seq_name: /cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-901-904-6667

seq_documentation_block:
; Sequence 6667, Application US/08901904
; GENERAL INFORMATION:
; APPLICANT: Gooding, Douglas H.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Naughton, Rebecca E.
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
; TITLE OF INVENTION: HUMAN PROSTATE
; NUMBER OF SEQUENCES: 6761
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/901,904
; FILING DATE: HEREWITH
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/023,288
; FILING DATE: JULY 29, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,302
; FILING DATE: DECEMBER 5, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PD-0198 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 6667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 1491822
US-08-901-904-6667

alignment_scores:
Quality: 109.00 Length: 22
Ratio: 4.955 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-215-435-225_COPY_1_22 x US-08-901-904-6667 ..

Align seg 1/1 to: US-08-901-904-6667 from: 1 to: 190

```
      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     54 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 103

     17 tMetValValThrGly 22
        ||||||||||||||||
    104 GATGGTGGTCACTGGA 119

seq_name: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7707

seq_documentation_block:
; Sequence 7707, Application US/09534853
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING SURFACE STRUCTURE AND MEMB
;  TITLE OF INVENTION:  MOLECULES
;  FILE REFERENCE: PD-1016 CIP
;  CURRENT APPLICATION NUMBER: US/09/534,853
;  CURRENT FILING DATE:  2000-03-24
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 28399
;  SOFTWARE: PERL Program
; SEQ ID NO 7707
;   LENGTH: 190
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00790232
US-09-534-853-7707


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-534-853-7707    ..

 Align seg 1/1  to: US-09-534-853-7707  from: 1  to: 190

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      54 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 103

      17 tMetValValThrGly 22
         ||||||||||||||||
     104 GATGGTGGTCACTGGA 119

 seq_name: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7685

 seq_documentation_block:
 ; Sequence 7685, Application US/09534853
 ; GENERAL INFORMATION:
 ;  APPLICANT: Seilhamer, Jeffrey J.
 ;  APPLICANT:  Delegeane, Angelo M.
 ;  APPLICANT:  Stuart, Susan G.
 ;  APPLICANT:  Stuve, Laura L.
 ;  APPLICANT:  Mullahy, Sara J.
 ;  APPLICANT:  Naughton, Rebecca E.
 ;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING SURFACE STRUCTURE AND MEMB
 ;  TITLE OF INVENTION:  MOLECULES
 ;  FILE REFERENCE: PD-1016 CIP
 ;  CURRENT APPLICATION NUMBER: US/09/534,853
 ;  CURRENT FILING DATE:  2000-03-24
 ;  Prior application data removed - consult PALM or file wrapper
 ;  NUMBER OF SEQ ID NOS: 28399
 ;  SOFTWARE: PERL Program
 ; SEQ ID NO 7685
 ;   LENGTH: 223
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: misc_feature
 ;   OTHER INFORMATION: Incyte ID No: hu00799304
 ;   NAME/KEY: unsure
 ;   LOCATION: 17
 ;   OTHER INFORMATION: a, t, c, g, or other
 US-09-534-853-7685


 alignment_scores:
             Quality:  109.00            Length:     22
               Ratio:  4.955              Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-09-215-435-225_COPY_1_22 x US-09-534-853-7685    ..

  Align seg 1/1  to: US-09-534-853-7685  from: 1  to: 223

        1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       64 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 113

       17 tMetValValThrGly 22
          ||||||||||||||||
      114 GATGGTGGTCACTGGA 129

 seq_name: /cgn1_7/ptodata/1/pna/US090_COMB.seq:US-09-035-171-1718

 seq_documentation_block:
 ; Sequence 1718, Application US/09035171
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Gooding, Douglas H.
 ;    APPLICANT:  Stuve, Laura L.
 ;    APPLICANT:  Stuart, Susan G.
 ;    APPLICANT:  Ito, Laura Y.
 ;    APPLICANT:  Akerblom, Ingrid E.
 ;    APPLICANT:  Delegeane, Angelo M.
 ;    APPLICANT:  Naughton, Rebecca E.
 ;    APPLICANT:  Klingler, Tod M.
 ;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
 ;    TITLE OF INVENTION:  LUNG TUMOR
 ;    NUMBER OF SEQUENCES:  2026
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
 ;      STREET:  3174 PORTER DRIVE
 ;      CITY:  PALO ALTO
 ;      STATE:  CALIFORNIA
 ;      COUNTRY:  USA
 ;      ZIP:  94304
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/09/035,171
 ;      FILING DATE:
 ;      CLASSIFICATION:
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  60/039,416
 ;      FILING DATE:
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  CERRONE, MICHAEL C.
 ;      REGISTRATION NUMBER:  39,132
 ;      REFERENCE/DOCKET NUMBER:  PD-0327P
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  (415) 855-0555
```

```
;       TELEFAX:  (415) 845-4166
;   INFORMATION FOR SEQ ID NO:  1718:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  239 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  2723986H1
US-09-035-171-1718


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-035-171-1718    ..

 Align seg 1/1  to: US-09-035-171-1718  from: 1  to: 239

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      10 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 59

      17 tMetValValThrGly 22
         ||||||||||||||||
      60 GATGGTGGTCACTGGA 75

seq_name: /cgn1_7/ptodata/1/pna/US095C_COMB.seq:US-09-540-212A-10721

seq_documentation_block:
; Sequence 10721, Application US/09540212A
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
;  FILE REFERENCE: PD-1034 CIP
;  CURRENT APPLICATION NUMBER: US/09/540,212A
;  CURRENT FILING DATE:  2000-03-31
;  NUMBER OF SEQ ID NOS: 67551
;  SOFTWARE: PERL Program
; SEQ ID NO 10721
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00358906
US-09-540-212A-10721


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-540-212A-10721    ..

 Align seg 1/1  to: US-09-540-212A-10721  from: 1  to: 239

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      10 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 59

      17 tMetValValThrGly 22
         ||||||||||||||||
      60 GATGGTGGTCACTGGA 75

seq_name: /cgn1_7/ptodata/1/pna/US6003_COMB.seq:US-60-039-416-1718

seq_documentation_block:
; Sequence 1718, Application US/60039416
; GENERAL INFORMATION:
;   APPLICANT:  Gooding, Douglas H.
;   APPLICANT:  Stuve, Laura L.
;   APPLICANT:  Stuart, Susan G.
;   APPLICANT:  Ito, Laura Y.
;   APPLICANT:  Akerblom, Ingrid E.
;   APPLICANT:  Delegeane, Angelo M.
;   APPLICANT:  Naughton, Rebecca E.
;   APPLICANT:  Klingler, Tod M.
;   TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;   TITLE OF INVENTION:  LUNG TUMOR
;   NUMBER OF SEQUENCES:  2026
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;     STREET:  3174 PORTER DRIVE
;     CITY:  PALO ALTO
;     STATE:  CALIFORNIA
;     COUNTRY:  USA
;     ZIP:  94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/60/039,416
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  CERRONE, MICHAEL C.
;     REGISTRATION NUMBER:  39,132
;     REFERENCE/DOCKET NUMBER:  PD-0327P
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 855-0555
;     TELEFAX:  (415) 845-4166
;  INFORMATION FOR SEQ ID NO:  1718:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  239 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   IMMEDIATE SOURCE:
;     CLONE:  2723986H1
US-60-039-416-1718


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-60-039-416-1718    ..

 Align seg 1/1  to: US-60-039-416-1718  from: 1  to: 239

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      10 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 59

      17 tMetValValThrGly 22
         ||||||||||||||||
      60 GATGGTGGTCACTGGA 75

seq_name: /cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-856-524-163
```

```
seq_documentation_block:
; Sequence 163, Application US/08856524
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  HUMAN FETAL LUNG
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/856,524
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LUTHER, BARBARA J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PD-0177P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 852-0195
;  INFORMATION FOR SEQ ID NO:  163:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1354794
US-08-856-524-163


alignment_scores:
            Quality:  109.00              Length:      22
              Ratio:  4.955                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-856-524-163    ..

 Align seg 1/1  to: US-08-856-524-163  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-856-524-905

seq_documentation_block:
; Sequence 905, Application US/08856524
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  HUMAN FETAL LUNG
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/856,524
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LUTHER, BARBARA J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PD-0177P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 852-0195
;  INFORMATION FOR SEQ ID NO:  905:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1357688
US-08-856-524-905


alignment_scores:
            Quality:  109.00              Length:      22
              Ratio:  4.955                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-856-524-905    ..

 Align seg 1/1  to: US-08-856-524-905  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-862-968-1249

seq_documentation_block:
; Sequence 1249, Application US/08862968
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
```

```
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  HUMAN PROSTATE
;    NUMBER OF SEQUENCES:  1716
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/862,968
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/019,396
;      FILING DATE:  MAY 24, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PD-0175 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 845-4166
;  INFORMATION FOR SEQ ID NO:  1249:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1347609
US-08-862-968-1249


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-862-968-1249    ..

 Align seg 1/1  to: US-08-862-968-1249  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      69 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 118

      17 tMetValValThrGly 22
         ||||||||||||||||
     119 GATGGTGGTCACTGGA 134

seq_name: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7694

seq_documentation_block:
; Sequence 7694, Application US/09534853
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING SURFACE STRUCTURE AND M
;  TITLE OF INVENTION:  MOLECULES
;  FILE REFERENCE: PD-1016 CIP
;  CURRENT APPLICATION NUMBER: US/09/534,853
;  CURRENT FILING DATE:  2000-03-24
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 28399
;  SOFTWARE: PERL Program
; SEQ ID NO 7694
;   LENGTH: 241
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00489356
;   NAME/KEY: unsure
;   LOCATION: 22
;   OTHER INFORMATION: a, t, c, g, or other
US-09-534-853-7694


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-534-853-7694    ..

 Align seg 1/1  to: US-09-534-853-7694  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      69 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 118

      17 tMetValValThrGly 22
         ||||||||||||||||
     119 GATGGTGGTCACTGGA 134

seq_name: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7700

seq_documentation_block:
; Sequence 7700, Application US/09534853
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING SURFACE STRUCTURE AND M
;  TITLE OF INVENTION:  MOLECULES
;  FILE REFERENCE: PD-1016 CIP
;  CURRENT APPLICATION NUMBER: US/09/534,853
;  CURRENT FILING DATE:  2000-03-24
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 28399
;  SOFTWARE: PERL Program
; SEQ ID NO 7700
;   LENGTH: 241
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00410317
;   NAME/KEY: unsure
;   LOCATION: 176
;   OTHER INFORMATION: a, t, c, g, or other
US-09-534-853-7700


alignment_scores:
```

```
           Quality:  109.00              Length:     22
             Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-534-853-7700   ..

 Align seg 1/1  to: US-09-534-853-7700  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7704

seq_documentation_block:
; Sequence 7704, Application US/09534853
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING SURFACE STRUCTURE AND MEME
;  TITLE OF INVENTION:  MOLECULES
;  FILE REFERENCE: PD-1016 CIP
;  CURRENT APPLICATION NUMBER: US/09/534,853
;  CURRENT FILING DATE:  2000-03-24
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 28399
;  SOFTWARE: PERL Program
; SEQ ID NO 7704
;   LENGTH: 241
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00156135
;   NAME/KEY: unsure
;   LOCATION: 129, 145, 152, 157, 160, 163, 170, 175, 181, 189, 194, 196, 200, 202, 205,
;   LOCATION: 218-219, 223, 225, 235, 238
;   OTHER INFORMATION: a, t, c, g, or other
US-09-534-853-7704


alignment_scores:
           Quality:  109.00              Length:     22
             Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-534-853-7704   ..

 Align seg 1/1  to: US-09-534-853-7704  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-017-766-163

seq_documentation_block:
; Sequence 163, Application US/60017766
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  HUMAN FETAL LUNG
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/017,766
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LUTHER, BARBARA J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PD-0177P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 852-0195
;  INFORMATION FOR SEQ ID NO:  163:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1354794
US-60-017-766-163


alignment_scores:
           Quality:  109.00              Length:     22
             Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-60-017-766-163   ..

 Align seg 1/1  to: US-60-017-766-163  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US6001_COMB.seq:US-60-017-766-905

seq_documentation_block:
; Sequence 905, Application US/60017766
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
```

```
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  HUMAN FETAL LUNG
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/017,766
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LUTHER, BARBARA J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PD-0177P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 852-0195
;  INFORMATION FOR SEQ ID NO:  905:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1357688
US-60-017-766-905


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-60-017-766-905   ..

 Align seg 1/1  to: US-60-017-766-905  from: 1  to: 241

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      47 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 96

      17 tMetValValThrGly 22
         ||||||||||||||||
      97 GATGGTGGTCACTGGA 112

seq_name: /cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-808-904-3732

seq_documentation_block:
; Sequence 3732, Application US/08808904
;  GENERAL INFORMATION:
;    APPLICANT:  Gooding, Douglas H.
;    APPLICANT:  Stuve, Laura L.
;    APPLICANT:  Stuart, Susan G.
;    APPLICANT:  Ito, Laura Y.
;    APPLICANT:  Akerblom, Ingrid E.
;    APPLICANT:  Delegeane, Angelo M.
;    APPLICANT:  Naughton, Rebecca E.
;    APPLICANT:  Klingler, Tod M.
;    TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM
;    TITLE OF INVENTION:  ASTHMATIC LUNG
;    NUMBER OF SEQUENCES:  5094
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,904
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/012,698
;      FILING DATE:  FEBRUARY 29, 1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/013,365
;      FILING DATE:  MARCH 13, 1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,122
;      FILING DATE:  SEPTEMBER 27, 1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CERRONE, MICHAEL C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PD-0120 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 855-0555
;      TELEFAX:  (415) 845-4166
;  INFORMATION FOR SEQ ID NO:  3732:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  244 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1985045
US-08-808-904-3732


alignment_scores:
            Quality:  109.00              Length:     22
              Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-08-808-904-3732   ..

 Align seg 1/1  to: US-08-808-904-3732  from: 1  to: 244

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      87 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 136

      17 tMetValValThrGly 22
         ||||||||||||||||
     137 GATGGTGGTCACTGGA 152
```



```
OM of: US-09-215-435-225_COPY_1_22 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:10 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8638 +   49.00  142.09  1.06
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-572 +   48.00  123.76  11.15     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4155 +   47.50  121.85  14.25
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-7131 +   47.00  135.59  2.45
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4631 +   45.50  119.26  19.87
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-8323 +   45.50  119.00  20.55
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-684-6 +   45.00  119.32  19.72     111
/cgn1_7/ptodata/1/pna/PCT_NEW_COMB.seq:PCT-US01-15332A-39 +   45.00  118.77  21.18     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-385-5 -   43.00  104.80  127.04    263
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-385-8 -   43.00  104.80  127.04    263
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-9 +   43.00  103.88  142.83    290
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-1 +   43.00  103.10  158.02    314
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-18 -   43.00  103.10  158.02    31
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-965 -   43.00  103.01  159.76    3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-214 +   43.00  102.21  177.11    3
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-1598 +   43.00  102.21  177.11
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-006-34 +   43.00   92.43  620.46    95
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-976-594-1083 -   42.50  100.44  222.12
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-689-908-381 -   42.00  125.75  8.65      2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-23150 -   42.00  122.32  13.42
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-1216 +   42.00  119.47  19.35
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-963-299-827 +   42.00  116.90  26.92     5
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1075 -   42.00  106.92  96.73
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-189-369 +   42.00  106.03  108.46    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-192-369 +   42.00  106.03  108.46    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-697-369 +   42.00  106.03  108.46    1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-824-369 +   42.00  106.03  108.46    1
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-324-050-15 +   42.00  102.61  168.24    24
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1086 +   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1087 -   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1088 -   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1089 -   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1090 +   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1091 +   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-1092 +   42.00   92.33  628.96
/cgn1_7/ptodata/1/pna/PCT_NEW_COMB.seq:PCT-US01-31303-1 +   42.00   84.97  1.6e+03   149
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-803-671B-3 -   42.00   70.89  9.6e+03   64
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-1868 -   41.50  118.77  21.17
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-2060 +   41.50  102.83  163.44
/cgn1_7/ptodata/1/pna/PCT_NEW_COMB.seq:PCT-US01-31303-2 +   41.00  119.58  19.08     300
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-9297 -   41.00  105.14  121.54
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-729 +   41.00  102.24  176.35
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-730 +   41.00  102.24  176.35
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-731 +   41.00  102.24  176.35
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1499 -   41.00   94.72  462.45

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-388-906A-8638

seq_documentation_block:
; Sequence 8638, Application US/09388906A
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael Andrew
;  TITLE OF INVENTION: Polynucleotides Isolated From Plants and
;  TITLE OF INVENTION:  Methods For Their Use
;  FILE REFERENCE: 11000.1013U
;  CURRENT APPLICATION NUMBER: US/09/388,906A
;  CURRENT FILING DATE: .1999-09-01
;  NUMBER OF SEQ ID NOS: 24843
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8638
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-388-906A-8638


alignment_scores:
            Quality:   49.00              Length:     21
              Ratio:  2.882                Gaps:      0
 Percent Similarity:  80.952   Percent Identity: 33.333

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-388-906A-8638   ..

 Align seg 1/1  to: US-09-388-906A-8638  from: 1  to: 377

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||:::|||::::::      :::|||   :::   :::::
     235 ATGGGATGGACTCAGAGAATTCTGAAGATGAGCCTAACGATGAGAGTTCT 284

      17 tMetValValThr 21
         ::::::::::|||
     285 ACTAATGATGACA 297

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-572

seq_documentation_block:
; Sequence 572, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 572
;   LENGTH: 1831
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-572


alignment_scores:
            Quality:   48.00              Length:     22
              Ratio:  3.429                Gaps:      0
 Percent Similarity:  63.636   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-60-325-448-572   ..

 Align seg 1/1  to: US-60-325-448-572  from: 1  to: 1831
```

```
      1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
        |||:::|||   |||      |||||||||   :::   |||||||||
    449 ATGGCCTGGCTCATGGCAAAGGTTACTGCAAATATGCATCTGGGGCTTGC 498

     17 tMetValValThrGly 22
           :::|||   |||
    499 TACATTGGTACATGGG 514

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4155

seq_documentation_block:
; Sequence 4155, Application US/09815242
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4155
;   LENGTH: 1902
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4155


alignment_scores:
            Quality:   47.50              Length:      21
              Ratio:  2.969                Gaps:       1
 Percent Similarity:  76.190   Percent Identity: 57.143

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-815-242-4155   ..

 Align seg 1/1  to: US-09-815-242-4155  from: 1  to: 1902

       3 Trp...ThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         |||   |||:::|||||||   :::|||:::::::|||||||||||||
    1443 TGGCGGACGCTTCGGCTTCGCAGTGCTGGCGTTCTTCTCGGTCTACCGCG 1492

      18 tValValThrGly 22
            ||||||   |||
    1493 AGTTGTTCGAGGT 1505

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-7131

 seq_documentation_block:
 ; Sequence 7131, Application US/09815242
 ; GENERAL INFORMATION:
 ;  APPLICANT: Haselbeck, Robert
 ;  APPLICANT:  Ohlsen, Kari L.
 ;  APPLICANT:  Zyskind, Judith W.
 ;  APPLICANT:  Wall, Daniel
 ;  APPLICANT:  Trawick, John D.
 ;  APPLICANT:  Carr, Grant J.
 ;  APPLICANT:  Yamamoto, Robert T.
 ;  APPLICANT:  Xu, H. Howard
 ;  TITLE OF INVENTION: Identification of Essential Genes in
 ;  TITLE OF INVENTION:  Prokaryotes
 ;  FILE REFERENCE: ELITRA.011A
 ;  CURRENT APPLICATION NUMBER: US/09/815,242
 ;  CURRENT FILING DATE:  2001-03-21
 ;  PRIOR APPLICATION NUMBER: 60/191,078
 ;  PRIOR FILING DATE: 2000-03-21
 ;  PRIOR APPLICATION NUMBER: 60/206,848
 ;  PRIOR FILING DATE: 2000-05-23
 ;  PRIOR APPLICATION NUMBER: 60/207,727
 ;  PRIOR FILING DATE: 2000-05-26
 ;  PRIOR APPLICATION NUMBER: 60/242,578
 ;  PRIOR FILING DATE: 2000-10-23
 ;  PRIOR APPLICATION NUMBER: 60/253,625
 ;  PRIOR FILING DATE: 2000-11-27
 ;  PRIOR APPLICATION NUMBER: 60/257,931
 ;  PRIOR FILING DATE: 2000-12-22
 ;  PRIOR APPLICATION NUMBER: 60/269,308
 ;  PRIOR FILING DATE: 2001-02-16
 ;  NUMBER OF SEQ ID NOS: 14110
 ;  SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 7131
 ;   LENGTH: 390
 ;   TYPE: DNA
 ;   ORGANISM: Haemophilus influenzae
 ;   FEATURE:
 ;   NAME/KEY: CDS
 ;   LOCATION: (1)...(390)
 US-09-815-242-7131


 alignment_scores:
             Quality:   47.00              Length:      18
               Ratio:  3.357                Gaps:       0
  Percent Similarity:  77.778   Percent Identity: 55.556

 alignment_block:
  US-09-215-435-225_COPY_1_22 x US-09-815-242-7131   ..

  Align seg 1/1  to: US-09-815-242-7131  from: 1  to: 390

        5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
          |||:::   :::|||   |||||||:::||||||     :::||||||||
       13 ATGAAAAAATTAACATTAGCATTGGTTTTAGGTTCAGCTTTAGTCGTGAC 62

       21 rGly 22
          ||||
       63 AGGT 66

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-4631

 seq_documentation_block:
 ; Sequence 4631, Application US/09815242
 ; GENERAL INFORMATION:
 ;  APPLICANT: Haselbeck, Robert
 ;  APPLICANT:  Ohlsen, Kari L.
 ;  APPLICANT:  Zyskind, Judith W.
 ;  APPLICANT:  Wall, Daniel
 ;  APPLICANT:  Trawick, John D.
 ;  APPLICANT:  Carr, Grant J.
 ;  APPLICANT:  Yamamoto, Robert T.
 ;  APPLICANT:  Xu, H. Howard
 ;  TITLE OF INVENTION: Identification of Essential Genes in
 ;  TITLE OF INVENTION:  Prokaryotes
 ;  FILE REFERENCE: ELITRA.011A
```

```
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4631
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4631


alignment_scores:
            Quality:   45.50              Length:     19
              Ratio:  2.844                Gaps:      1
 Percent Similarity:  84.211   Percent Identity: 52.632

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-815-242-4631   ..

 Align seg 1/1  to: US-09-815-242-4631  from: 1  to: 1311

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||   :::   :::|||   :::||||||:::||||||||||:::|| 
     415 GGCTGG...CTAACGATTGTTATCGGCGCACTCGTATTAGGATTAATCAT 461

      18 tValVal 20
         |:::|||
     462 GGCAGTC 468

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-8323

seq_documentation_block:
; Sequence 8323, Application US/09815242
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8323
;   LENGTH: 1347
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1347)
US-09-815-242-8323


alignment_scores:
            Quality:   45.50              Length:     19
              Ratio:  2.844                Gaps:      1
 Percent Similarity:  84.211   Percent Identity: 52.632

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-815-242-8323   ..

 Align seg 1/1  to: US-09-815-242-8323  from: 1  to: 1347

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||   :::   :::|||   :::||||||:::||||||||||:::|| 
     433 GGCTGG...CTAACGATTGTTATCGGCGCACTCGTATTAGGATTAATCAT 479

      18 tValVal 20
         |:::|||
     480 GGCAGTC 486

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-974-684-6

seq_documentation_block:
; Sequence 6, Application US/09974684
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: Human Nucleic Acid Sequences Encoding
;  TITLE OF INVENTION:  G-Protein Coupled Receptor-like Homologs and Uses Therefor
;  FILE REFERENCE: 35800/239478
;  CURRENT APPLICATION NUMBER: US/09/974,684
;  CURRENT FILING DATE:  2001-10-10
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 1110
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-974-684-6


alignment_scores:
            Quality:   45.00              Length:     23
              Ratio:  2.647                Gaps:      1
 Percent Similarity:  73.913   Percent Identity: 47.826

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-974-684-6   ..

 Align seg 1/1  to: US-09-974-684-6  from: 1  to: 1110

       2 GlyTrp......ThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLe 16
         ||||||      |||:::   ::::::||||||||||   |||:::   || 
     649 GGTTGGTGTCAGACATTTGATTTCATCACTGCAGCGTGGCTGATTTTTTT 698

      16 uMetMetValValThrGly 22
         |:::|||||||:::   |||
     699 ATTCATGGTTCTCTGTGGG 717

seq_name: /cgn1_7/ptodata/1/pna/PCT_NEW_COMB.seq:PCT-US01-15332A-39

seq_documentation_block:
```

```
; Sequence 39, Application PC/TUS0115332A
; GENERAL INFORMATION:
;  APPLICANT: Brown, Joseph P.
;  APPLICANT:  Miller, Margaret
;  APPLICANT:  Burmer, Glenna
;  APPLICANT:  Fabre-Suver, Christine
;  APPLICANT:  Pritchard, David
;  APPLICANT:  LifeSpan Biosciences, Inc.
;  TITLE OF INVENTION: Nucleic Acid Sequences for Novel GPCRs
;  FILE REFERENCE: 017473-003710PC
;  CURRENT APPLICATION NUMBER: PCT/US01/15332A
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: 60/203,217
;  PRIOR FILING DATE: 2000-05-11
;  PRIOR APPLICATION NUMBER: 60/205,945
;  PRIOR FILING DATE: 2000-05-18
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 1176
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Ls191218 cluster
PCT-US01-15332A-39


alignment_scores:
            Quality:   45.00              Length:     23
              Ratio:  2.647                Gaps:      1
 Percent Similarity:  73.913   Percent Identity: 47.826

alignment_block:
 US-09-215-435-225_COPY_1_22 x PCT-US01-15332A-39    ..

 Align seg 1/1  to: PCT-US01-15332A-39  from: 1  to: 1176

       2 GlyTrp......ThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLe 16
         ||||||      |||:::   ::::::|||||||||   |||:::   ||
     680 GGTTGGTGTCAGACATTTGATTTCATCACTGCAGCGTGGCTGATTTTTTT 729

      16 uMetMetValValThrGly 22
         |:::|||||||:::   |||
     730 ATTCATGGTTCTCTGTGGG 748

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-385-5

seq_documentation_block:
; Sequence 5, Application US/09978385
; GENERAL INFORMATION:
;  APPLICANT: Piddington, Christopher S.
;  APPLICANT:  Petrie, Charles
;  APPLICANT:  Shoemaker, Kimberly E.
;  APPLICANT:  Bishop, Paul D.
;  TITLE OF INVENTION: ZACE2: A HUMAN METALLOENZYME
;  FILE REFERENCE: 99-24C1
;  CURRENT APPLICATION NUMBER: US/09/978,385
;  CURRENT FILING DATE:  2001-10-16
;  PRIOR APPLICATION NUMBER: 60/133,952
;  PRIOR FILING DATE: 1999-05-13
;  PRIOR APPLICATION NUMBER: 60/151,181
;  PRIOR FILING DATE: 1999-08-27
;  PRIOR APPLICATION NUMBER: 09/563,516
;  PRIOR FILING DATE: 2000-05-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 2638
;   TYPE: DNA
;   ORGANISM: Mouse
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (106)...(2520)
US-09-978-385-5


alignment_scores:
            Quality:   43.00              Length:     21
              Ratio:  2.688 .              Gaps:      0
 Percent Similarity:  76.190   Percent Identity: 38.095

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-978-385-5/rev    ..

 Align seg 1/1  to reverse of: US-09-978-385-5  from: 1  to: 2638

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ::::::::::|||   |||   :::|||:::|||||||||   |||   ::
    2559 TTAGCATTTACATACAGGAAGATGACAAGTGCTTTGCTAAAAGGAAGTCT 2510

      17 tMetValValThr 21
         :   ::::::|||
    2509 GAGCATCATCACT 2497

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-978-385-8

seq_documentation_block:
; Sequence 8, Application US/09978385
; GENERAL INFORMATION:
;  APPLICANT: Piddington, Christopher S.
;  APPLICANT:  Petrie, Charles
;  APPLICANT:  Shoemaker, Kimberly E.
;  APPLICANT:  Bishop, Paul D.
;  TITLE OF INVENTION: ZACE2: A HUMAN METALLOENZYME
;  FILE REFERENCE: 99-24C1
;  CURRENT APPLICATION NUMBER: US/09/978,385
;  CURRENT FILING DATE:  2001-10-16
;  PRIOR APPLICATION NUMBER: 60/133,952
;  PRIOR FILING DATE: 1999-05-13
;  PRIOR APPLICATION NUMBER: 60/151,181
;  PRIOR FILING DATE: 1999-08-27
;  PRIOR APPLICATION NUMBER: 09/563,516
;  PRIOR FILING DATE: 2000-05-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 2638
;   TYPE: DNA
;   ORGANISM: Mouse
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (106)...(2520)
US-09-978-385-8


alignment_scores:
            Quality:   43.00              Length:     21
              Ratio:  2.688                Gaps:      0
 Percent Similarity:  76.190   Percent Identity: 38.095

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-978-385-8/rev    ..

 Align seg 1/1  to reverse of: US-09-978-385-8  from: 1  to: 2638

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ::::::::::|||   |||   :::|||:::|||||||||   |||   ::
    2559 TTAGCATTTACATACAGGAAGATGACAAGTGCTTTGCTAAAAGGAAGTCT 2510

      17 tMetValValThr 21
         :   ::::::|||
    2509 GAGCATCATCACT 2497

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-9

seq_documentation_block:
```

```
; Sequence 9, Application US/09644600
; GENERAL INFORMATION:
;  APPLICANT: O'Brien, Timothy J.
;  APPLICANT:  Tanimoto, Hirotoshi
;  TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
;  TITLE OF INVENTION:  Overexpressed in Carcinomas
;  FILE REFERENCE: D6064CIP/D
;  CURRENT APPLICATION NUMBER: US/09/644,600
;  CURRENT FILING DATE:  2000-08-23
;  PRIOR APPLICATION NUMBER: 09/421,213
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: 09/027,337
;  PRIOR FILING DATE: 1998-02-20
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 9
;   LENGTH: 2900
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: SNC-19; GeneBank Accession No. #U20428
US-09-644-600-9


alignment_scores:
            Quality:   43.00              Length:     16
              Ratio:  3.071                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-644-600-9   ..

 Align seg 1/1  to: US-09-644-600-9  from: 1  to: 2900

       7 LeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValThrGly 22
         :::|||   ||||||:::|||:::|||||||::::::|||:::   |||
       7 GTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGG 54

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-1

seq_documentation_block:
; Sequence 1, Application US/09644600
; GENERAL INFORMATION:
;  APPLICANT: O'Brien, Timothy J.
;  APPLICANT:  Tanimoto, Hirotoshi
;  TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
;  TITLE OF INVENTION:  Overexpressed in Carcinomas
;  FILE REFERENCE: D6064CIP/D
;  CURRENT APPLICATION NUMBER: US/09/644,600
;  CURRENT FILING DATE:  2000-08-23
;  PRIOR APPLICATION NUMBER: 09/421,213
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: 09/027,337
;  PRIOR FILING DATE: 1998-02-20
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: TADG-15
US-09-644-600-1


alignment_scores:
            Quality:   43.00              Length:     16
              Ratio:  3.071                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-644-600-1   ..

 Align seg 1/1  to: US-09-644-600-1  from: 1  to: 3147

       7 LeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValThrGly 22
         :::|||   ||||||:::|||:::|||||||::::::|||:::   |||
     188 GTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGG 235

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-644-600-18

seq_documentation_block:
; Sequence 18, Application US/09644600
; GENERAL INFORMATION:
;  APPLICANT: O'Brien, Timothy J.
;  APPLICANT:  Tanimoto, Hirotoshi
;  TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
;  TITLE OF INVENTION:  Overexpressed in Carcinomas
;  FILE REFERENCE: D6064CIP/D
;  CURRENT APPLICATION NUMBER: US/09/644,600
;  CURRENT FILING DATE:  2000-08-23
;  PRIOR APPLICATION NUMBER: 09/421,213
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: 09/027,337
;  PRIOR FILING DATE: 1998-02-20
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
;   LENGTH: 3147
;   TYPE: RNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Antisense of TADG-15
US-09-644-600-18


alignment_scores:
            Quality:   43.00              Length:     16
              Ratio:  3.071                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 50.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-644-600-18/rev   ..

 Align seg 1/1  to reverse of: US-09-644-600-18  from: 1  to: 3147

       7 LeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValThrGly 22
         :::|||   ||||||:::|||:::|||||||::::::|||:::   |||
    2960 GTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGG 2913

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-965

seq_documentation_block:
; Sequence 965, Application US/09954456
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
```

```
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 965
;   LENGTH: 3175
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-965


alignment_scores:
            Quality:   43.00            Length:     21
              Ratio:  3.071              Gaps:      0
 Percent Similarity:  66.667   Percent Identity: 38.095

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-954-456-965/rev    ..

 Align seg 1/1  to reverse of: US-09-954-456-965  from: 1  to: 3175

       2 GlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMe 18
         ||||||||||:::   |||   :::   |||      |||  ::::::::
    2039 GGCTGGACTGTCTCACTTTGTTCTGAGGCTGCTCCTTTGAAAATCCTGTT 1990

      18 tValValThrGly 22
         ::::|||   |||
    1989 AATGGTGAGAGGC 1977

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-954-456-214

seq_documentation_block:
; Sequence 214, Application US/09954456
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 214
;   LENGTH: 3451
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-214


 alignment_scores:
             Quality:   43.00            Length:     21
               Ratio:  2.529              Gaps:      0
  Percent Similarity:  80.952   Percent Identity: 38.095

alignment_block:
 US-09-215-435-225_COPY_1_22 x US-09-954-456-214    ..

 Align seg 1/1  to: US-09-954-456-214  from: 1  to: 3451

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||   :::|||   |||::::::   |||::::::::::|||   ::
    1693 ATGGGCATCGTCATGTCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGT 1742

      17 tMetValValThr 21
         ::::|||:::|||
    1743 GCTGGTCATCACA 1755
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225_COPY_1_22 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:31 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_1_22
Query length: 22
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est93:BF853580         +  109.00  319.98  1.1e-08   242  ! BF853580 MR2-EN0093-211200-002-
gb_est93:BF853537         +  109.00  315.14  2.0e-08   407  ! BF853537 MR2-EN0093-211200-002-
gb_est93:BF853920         +  109.00  313.25  2.5e-08   499  ! BF853920 MR2-EN0093-261200-004-
gb_est93:BF853928         +  109.00  312.59  2.7e-08   536  ! BF853928 MR2-EN0093-261200-004-
gb_est93:BF853000         +  109.00  312.21  2.8e-08   558  ! BF853000 MR2-EN0093-191200-001-
gb_est93:BF852995         +  109.00  312.13  2.9e-08   563  ! BF852995 MR2-EN0093-191200-001-
gb_est93:BF852989         +  109.00  311.49  3.1e-08   603  ! BF852989 MR2-EN0093-191200-001-
gb_est102:BG545668        +  109.00  309.47  4.0e-08   749  ! BG545668 602572933F1 NIH_MGC_77
gb_est45:AW327171         +   95.00  274.79  3.5e-06   352  ! AW327171 20827 MARC 1BOV Bos ta
gb_est93:BF890431         +   95.00  272.42  4.7e-06   454  ! BF890431 292000 MARC 3BOV Bos t
gb_est105:H83784          +   82.00  233.98  0.0006    440  ! H83784 ys64b02.r1 Soares retina
gb_est93:BF846063         -   82.00  233.49  0.0007    464  ! BF846063 QV0-EN0055-181000-439-
gb_est93:BF846065         -   82.00  233.43  0.0007    467  ! BF846065 QV0-EN0055-181000-439-
gb_est90:BF599823         +   68.00  192.86  0.1266    413  ! BF599823 264147 MARC 3BOV Bos t
gb_est93:BF848191         +   64.50  181.37  0.5526    463  ! BF848191 QV0-EN0055-031100-480-
gb_est93:BF848198         +   61.50  172.44  1.74      463  ! BF848198 QV0-EN0055-031100-480-
gb_est81:BE934438         +   57.00  165.74  4.10      225  ! BE934438 MR0-HT0923-270800-202-
gb_est101:BG483879        +   56.50  152.59  22.18     789  ! BG483879 602504455F1 NIH_MGC_77
gb_est41:AW045423         +   55.00  155.54  15.18     355  ! AW045423 UI-M-BH1-akr-a-08-0-UI
gb_est93:BF846059         +   55.00  153.09  20.78     462  ! BF846059 QV0-EN0055-181000-439-
gb_est45:AW277752         -   55.00  152.81  21.54     476  ! AW277752 sf86b07.y1 Gm-c1019 Gl
gb_est25:AI853718         +   55.00  152.78  21.64     478  ! AI853718 UI-M-BH0-ajq-d-11-0-UI
gb_est80:BE863333         -   55.00  152.38  22.78     499  ! BE863333 UI-M-BH0-ajq-d-11-0-UI
gb_est49:AW620538         +   55.00  152.14  23.49     512  ! AW620538 sj06d11.y1 Gm-c1032 Gl
gb_est46:AW396613         +   55.00  151.92  24.15     524  ! AW396613 sg80c10.y1 Gm-c1026 Gl
gb_est4:AA239366          -   55.00  151.47  25.59     550  ! AA239366 my38d03.r1 Barstead mc
gb_est30:AU080994         -   55.00  150.98  27.26     580  ! AU080994 AU080994 Sugano mouse
gb_est110:W63852          -   55.00  150.95  27.37     582  ! W63852 md85d01.r1 Soares mouse
gb_est51:AW742322         -   55.00  149.76  31.86     661  ! AW742322 up55e10.y1 Soares_mous
gb_est72:BE226513         -   55.00  149.53  32.84     678  ! BE226513 ia21g02.y1 Mouse E10 5
gb_est30:AU080413         -   55.00  148.11  39.40     790  ! AU080413 AU080413 Sugano mouse
gb_est30:AU080863         -   55.00  147.71  41.43     824  ! AU080863 AU080863 Sugano mouse
gb_est30:AU051563         -   55.00  147.27  43.84     864  ! AU051563 AU051563 Sugano mouse
gb_est18:AI322707         -   55.00  146.97  45.60     893  ! AI322707 mr23a12.y1 Soares mous
gb_est98:BG296698         -   55.00  146.90  45.97     899  ! BG296698 602393647F1 NIH_MGC_94
gb_est30:AU051486         -   55.00  146.70  47.19     919  ! AU051486 AU051486 Sugano mouse
gb_htc:AK005078           -   55.00  141.52  91.68    1604  ! AK005078 Mus musculus adult ma
gb_htc:AK014392           +   55.00  136.65  171.33   2710  ! AK014392 Mus musculus adult ma
gb_est47:AW477165         +   54.50  148.34  38.22     656  ! AW477165 ga42f11.y1 Moss EST li
gb_est28:AL042119         -   53.50  146.74  46.96     566  ! AL042119 DKFZp434B0820_r1 434 (
gb_gss9:AQ284999          -   53.00  150.32  29.66     328  ! AQ284999 RPCI11-87D16.TV RPCI-1
gb_est99:BG354108         +   52.00  142.67  79.11     542  ! BG354108 00615 leafy spurge
gb_est93:BF848190         +   51.50  142.65  79.37     463  ! BF848190 QV0-EN0055-031100-4
gb_est108:T20428          -   51.00  142.99  75.91     380  ! T20428 2436 Lambda-PRL2 Arab
gb_est106:L33500          +   51.00  142.68  79.02     393  ! L33500 BNAESTF044 Mustard fl

seq_name: gb_est93:BF853580

seq_documentation_block:
LOCUS       BF853580      242 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-211200-002-b03 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853580
VERSION     BF853580.1  GI:12241324
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            211200-002-b03&t3=2000-12-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 242.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       44 a     64 c     74 g     60 t
ORIGIN



alignment_scores:
           Quality:  109.00              Length:     22
             Ratio:  4.955                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF853580   ..

 Align seg 1/1  to: BF853580  from: 1  to: 242

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140
```

```
     17 tMetValValThrGly 22
        ||||||||||||||||
    141 GATGGTGGTCACTGGA 156

seq_name: gb_est93:BF853537

seq_documentation_block:
LOCUS       BF853537      407 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-211200-002-h10 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853537
VERSION     BF853537.1  GI:12241281
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            211200-002-h10&t3=2000-12-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       87 a    104 c    124 g     92 t
ORIGIN


alignment_scores:
           Quality:  109.00            Length:     22
             Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF853537   ..

 Align seg 1/1  to: BF853537  from: 1  to: 407

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGly 22
```

```
         ||||||||||||||||
     141 GATGGTGGTCACTGGA 156

seq_name: gb_est93:BF853920

seq_documentation_block:
LOCUS       BF853920      499 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-d01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853920
VERSION     BF853920.1  GI:12241664
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            261200-004-d01&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 227.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      119 a    129 c    148 g    103 t
ORIGIN


alignment_scores:
           Quality:  109.00            Length:     22
             Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF853920   ..

 Align seg 1/1  to: BF853920  from: 1  to: 499

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      48 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 97

      17 tMetValValThrGly 22
```

```
       ||||||||||||||||
    98 GATGGTGGTCACTGGA 113

seq_name: gb_est93:BF853928

seq_documentation_block:
LOCUS       BF853928      536 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-g12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853928
VERSION     BF853928.1  GI:12241672
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            261200-004-g12&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      121 a    139 c    164 g    112 t
ORIGIN


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF853928   ..

 Align seg 1/1  to: BF853928  from: 1  to: 536

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGly 22
         ||||||||||||||||
```

```
     141 GATGGTGGTCACTGGA 156

seq_name: gb_est93:BF853000

seq_documentation_block:
LOCUS       BF853000      558 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-f12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853000
VERSION     BF853000.1  GI:12240848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-f12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 558.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      124 a    146 c    163 g    124 t      1 others
ORIGIN


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF853000   ..

 Align seg 1/1  to: BF853000  from: 1  to: 558

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      93 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 142

      17 tMetValValThrGly 22
         ||||||||||||||||
     143 GATGGTGGTCACTGGA 158
```

```
seq_name: gb_est93:BF852995

seq_documentation_block:
LOCUS       BF852995      563 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-e01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852995
VERSION     BF852995.1  GI:12240843
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-e01&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      126 a    148 c    164 g    125 t
ORIGIN


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF852995   ..

 Align seg 1/1  to: BF852995  from: 1  to: 563

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGly 22
         ||||||||||||||||
     141 GATGGTGGTCACTGGA 156
```

```
seq_name: gb_est93:BF852989

seq_documentation_block:
LOCUS       BF852989      603 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-a12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852989
VERSION     BF852989.1  GI:12240837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-a12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 560.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      140 a    157 c    171 g    135 t
ORIGIN


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF852989   ..

 Align seg 1/1  to: BF852989  from: 1  to: 603

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 ATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCAT 140

      17 tMetValValThrGly 22
         ||||||||||||||||
     141 GATGGTGGTCACTGGA 156

seq_name: gb_est102:BG545668
```

```
seq_documentation_block:
LOCUS       BG545668      749 bp    mRNA            EST       04-APR-2001
DEFINITION  602572933F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4700840 5',
            mRNA sequence.
ACCESSION   BG545668
VERSION     BG545668.1  GI:13544333
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1535  row: b  column: 09
            High quality sequence stop: 740.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4700840"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      184 a    206 c    205 g    154 t
ORIGIN


alignment_scores:
            Quality:  109.00            Length:     22
              Ratio:  4.955               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BG545668   ..

 Align seg 1/1  to: BG545668  from: 1  to: 749

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      86 ATGGGTTGGACAATGAGGCTGGTTACAGCAGCACTGTTACTGGGTCTCAT 135

      17 tMetValValThrGly 22
         ||||||||||||||||
     136 GATGGTGGTCACTGGA 151

seq_name: gb_est45:AW327171

seq_documentation_block:
LOCUS       AW327171      352 bp    mRNA            EST       10-JUL-2000
DEFINITION  20827 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW327171
VERSION     AW327171.1  GI:6763092
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 13  row: O  column: 22
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT       65 a    107 c    107 g     73 t
ORIGIN


alignment_scores:
            Quality:   95.00            Length:     22
              Ratio:  4.524               Gaps:      0
 Percent Similarity:  95.455   Percent Identity: 86.364

alignment_block:
 US-09-215-435-225_COPY_1_22 x AW327171   ..

 Align seg 1/1  to: AW327171  from: 1  to: 352

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     108 ATGGGCTGGACCATGAGGCTGGTCACAGCAGCCCTGCTCCTGGGCCTCGC 157

      17 tMetValValThrGly 22
         ::::::||||||||||
     158 GGTGGCGGTCACTGGA 173

seq_name: gb_est93:BF890431

seq_documentation_block:
LOCUS       BF890431      454 bp    mRNA            EST       18-JAN-2001
DEFINITION  292000 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF890431
VERSION     BF890431.1  GI:12281889
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 454)
```

```
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 57  row: E  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       93 a    134 c    132 g     94 t      1 others
ORIGIN


alignment_scores:
            Quality:   95.00             Length:     22
              Ratio:  4.524               Gaps:      0
 Percent Similarity:  95.455   Percent Identity: 86.364

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF890431   ..

 Align seg 1/1  to: BF890431  from: 1  to: 454

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      74 ATGGGCTGGACCATGAGGCTGGTCACAGCAGCCCTGCTCCTGGGCCTCGC 123

      17 tMetValValThrGly 22
         ::::::|||||||||
     124 GGTGGCGGTCACTGGA 139

seq_name: gb_est105:H83784

seq_documentation_block:
LOCUS        H83784         440 bp    mRNA            EST       13-NOV-1995
DEFINITION   ys64b02.r1 Soares retina N2b4HR Homo sapiens cDNA clone
             IMAGE:219531 5', mRNA sequence.
ACCESSION    H83784
VERSION      H83784.1  GI:1062455
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 440)
  AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
             ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
             Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
             ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE      The WashU-Merck EST Project
  JOURNAL    Unpublished (1995)
COMMENT      Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             High quality sequence stops: 364
             Source: IMAGE Consortium, LLNL
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 911   Std Error: 0.00
             Seq primer: M13RP1
             High quality sequence stop: 364.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Homo sapiens"
                     /db_xref="GDB:3847524"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:219531"
                     /clone_lib="Soares retina N2b4HR"
                     /sex="male"
                     /tissue_type="retina"
                     /dev_stage="55 year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). The retinas were obtained from a 55 year old
                     Caucasian  and total cellular poly(A)+ RNA was extracted 6
                     hrs after their removal. The retina RNA  was kindly
                     provided by Roderick R. McInnes M.D. Ph.D. from the
                     University  of Toronto. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. "
BASE COUNT       99 a     99 c    131 g    103 t      8 others
ORIGIN


alignment_scores:
            Quality:   82.00             Length:     22
              Ratio:  3.905               Gaps:      0
 Percent Similarity:  95.455   Percent Identity: 95.455

alignment_block:
 US-09-215-435-225_COPY_1_22 x H83784   ..

 Align seg 1/1  to: H83784  from: 1  to: 440

       1 MetGlyTrpThrMetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMe 17
         |||||||||||||||||||||||||   ||||||||||||||||||||
      44 ATGGGTTGGACAATGAGGCTGGTCACACA.GCACTGTTACTGGGTCTCAT 92

      17 tMetValValThrGly 22
         ||||||||||||||||
      93 GATGGTGGTCACTGGA 108

seq_name: gb_est93:BF846063

seq_documentation_block:
LOCUS        BF846063       464 bp    mRNA            EST       16-JAN-2001
DEFINITION   QV0-EN0055-181000-439-f11 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF846063
VERSION      BF846063.1  GI:12233213
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 464)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
```

```
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-EN0055-
            181000-439-f11&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       94 a    134 c    122 g    114 t
ORIGIN


alignment_scores:
            Quality:   82.00             Length:     18
              Ratio:  4.556               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF846063/rev   ..

 Align seg 1/1  to reverse of: BF846063  from: 1  to: 464

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     444 ATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCATGATGGTGGTCAC 395

      21 rGly 22
         ||||
     394 TGGA 391

seq_name: gb_est93:BF846065

seq_documentation_block:
LOCUS       BF846065      467 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-181000-439-g06 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846065
VERSION     BF846065.1  GI:12233215
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-EN0055-
            181000-439-g06&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 467.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       94 a    134 c    126 g    113 t
ORIGIN


alignment_scores:
            Quality:   82.00             Length:     18
              Ratio:  4.556               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF846065/rev   ..

 Align seg 1/1  to reverse of: BF846065  from: 1  to: 467

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     447 ATGAGGCTGGTCACAGCAGCACTGTTACTGGGTCTCATGATGGTGGTCAC 398

      21 rGly 22
         ||||
     397 TGGA 394

seq_name: gb_est90:BF599823

seq_documentation_block:
LOCUS       BF599823      413 bp    mRNA            EST       13-DEC-2000
DEFINITION  264147 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF599823
VERSION     BF599823.1  GI:11696691
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 413)
```

```
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 35  row: D  column: 18
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       87 a    126 c    107 g     93 t
ORIGIN


alignment_scores:
            Quality:   68.00            Length:     18
              Ratio:  4.000              Gaps:      0
 Percent Similarity:  94.444   Percent Identity: 83.333

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF599823   ..

 Align seg 1/1  to: BF599823  from: 1  to: 413

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         |||||||||||||||||||||||||||||||||||||   ::::::|||||
       2 ATGAGGCTGGTCACAGCAGCCCTGCTCCTGGGCCTCGCGGTGGCGGTCAC 51

      21 rGly 22
         ||||
      52 TGGA 55

seq_name: gb_est93:BF848191

seq_documentation_block:
LOCUS       BF848191      463 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-031100-480-c10 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF848191
VERSION     BF848191.1  GI:12235341
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-EN0055-
            031100-480-c10&t3=2000-11-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0055"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      116 a    119 c    136 g     92 t
ORIGIN


alignment_scores:
            Quality:   64.50            Length:     18
              Ratio:  3.794              Gaps:      1
 Percent Similarity:  94.444   Percent Identity: 88.889

alignment_block:
 US-09-215-435-225_COPY_1_22 x BF848191   ..

 Align seg 1/1  to: BF848191  from: 1  to: 463

       5 MetArgLeuValThrAlaAlaLeuLeuLeuGlyLeuMetMetValValTh 21
         ||||||||||||||||||||   :::|||||||||||||||||||||||||
      30 ATGAGGCTGGTCACAGCAGCA...GTACTGGGTCTCATGATGGTGGTCAC 76

      21 rGly 22
         ||||
      77 TGGA 80
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225_COPY_23_227 to: GenEmbl:*   out_format : pfs

Date: Nov 15, 2001 12:52 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2129/app_query.fasta_1.1519
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_11006 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 7698.940000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pat2:AX083412       +  1115.00 2095.98  1.6e-108  681   ! AX083412 Sequence 104 from Pate
gb_pat2:AX083422       +  1115.00 2093.47  2.3e-108  897   ! AX083422 Sequence 114 from Pate
gb_pat1:AX060293       +  1111.00 2085.55  6.3e-108  932   ! AX060293 Sequence 1 from Patent
gb_sts2:G27363         -   444.00  833.20  3.6e-38   352   ! G27363 human STS SHGC-31033, se
gb_htg13:AC055854      -   262.50  433.10  6.9e-16   179814  ! AC055854 Homo sapiens chromo
gb_ro2:RNPBP           +   237.50  432.87  7.1e-16   1037  ! X71873 R.norvegicus mRNA for p
gb_pat2:E05646         +   237.50  432.79  7.2e-16   1047  ! E05646 cDNA encoding rat hippo
gb_ro2:RNPEABP         +   237.50  432.55  7.4e-16   1075  ! X75253 R.norvegicus phosphatid
gb_htg12:AC037441      -   236.00  383.90  3.8e-13   162795  ! AC037441 Homo sapiens clone
gb_htg17:AC084081      -   236.00  383.53  4.0e-13   169645  ! AC084081 Homo sapiens chromo
gb_in4:OVD1PROTN       +   234.50  432.03  7.9e-16   611   ! X87991 O.volvulus mRNA for d1 p
gb_in3:CELF40A3        +   230.00  385.10  3.3e-13   41155  ! AF016423 Caenorhabditis elega
gb_htg1:AC006915       +   230.00  374.33  1.3e-12   134095  ! AC006915 Caenorhabditis eleg
gb_htg4:AC012978       -   229.00  379.25  6.9e-13   63551  ! AC012978 Drosophila melanogas
gb_in1:AC005891        +   229.00  372.24  1.7e-12   136933  ! AC005891 Drosophila melanoga
gb_in1:AE003637        +   229.00  367.77  3.0e-12   223646  ! AE003637 Drosophila melanoga
gb_ro1:AF226629        +   227.50  412.60  9.6e-15   1205  ! AF226629 Rattus norvegicus pho
gb_ro1:AF300422        +   225.00  414.80  7.2e-15   564   ! AF300422 Mus musculus phosphati
gb_ro1:AB046417        +   225.00  407.89  1.8e-14   1202  ! AB046417 Mus musculus HCNPpp m
gb_htg6:AC020316       -   222.50  364.56  4.5e-12   82631  ! AC020316 Drosophila melanogas
gb_in2:AE003675        +   222.50  359.38  8.8e-12   145786  ! AE003675 Drosophila melanoga
gb_in4:OVD2PROTN       +   219.00  404.30  2.8e-14   514   ! X87990 O.volvulus mRNA for d1 p
gb_ro1:MMU43206        +   219.00  396.71  7.3e-14   1180  ! U43206 Mus musculus phosphatid
gb_htg5:AC016445       +   207.00  334.22  2.2e-10   92569  ! AC016445 Drosophila melanogas
gb_pl2:AF332406        +   206.50  380.52  5.9e-13   522   ! AF332406 Arabidopsis thaliana p
gb_pl2:AF147721        +   206.50  376.30  1.0e-12   829   ! AF147721 Arabidopsis thaliana E
gb_pr10:MFPBP          +   206.00  370.29  2.2e-12   1444  ! X73137 M.fascicularis mRNA for
em_pat:E09005          +   205.00  376.98  9.2e-13   564   ! E09005 Human cDNA encoding a ph
gb_pr9:HSRNAPBP        +   205.00  376.88  9.4e-13   570   ! X85033 H.sapiens mRNA for phosp
em_pat:E09006          +   205.00  371.74  1.8e-12   1001  ! E09006 Human cDNA encoding a p
gb_pr10:HUMHRPBP       +   205.00  368.46  2.8e-12   1434  ! D16111 Human mRNA for human hn
gb_pr9:HSPEABP         +   205.00  368.40  2.8e-12   1444  ! X75252 H.sapiens phosphatidyle
gb_pat2:E05647         +   205.00  368.38  2.8e-12   1447  ! E05647 cDNA encoding human hig
gb_pr10:S76773         +   202.00  369.04  2.6e-12   723   ! S76773 neuropolypeptide h3 [hum
gb_pat2:I11816         +   201.50  366.92  3.4e-12   822   ! I11816 Sequence 1 from Patent U
gb_in4:ONGANTIG        +   201.50  366.92  3.4e-12   822   ! M27807 O.volvulus O.V.-16 antig
gb_in1:AB035321        +   196.00  360.46  7.7e-12   534   ! AB035321 Dirofilaria immitis mR
gb_htg13:AC062037      -   194.00  303.20  1.2e-08   187355  ! AC062037 Homo sapiens chromo
gb_htg12:AC037441      +   191.50  299.76  1.8e-08   162795  ! AC037441 Homo sapiens clone
gb_htg5:AC016445       -   190.50  303.02  1.2e-08   92569  ! AC016445 Drosophila melanogas
gb_pl2:AF152907        +   182.00  334.09  2.3e-10   528   ! AF152907 Arabidopsis thaliana t
gb_pl1:AB027506        +   182.00  330.31  3.7e-10   799   ! AB027506 Arabidopsis thalian
gb_htg4:AC013131       +   182.00  297.12  2.6e-08   30334  ! AC013131 Drosophila melano
gb_in1:AC004439        +   182.00  287.82  8.5e-08   84055  ! AC004439 Drosophila melano
gb_in1:AE003586        +   182.00  276.02  3.9e-07   306352  ! AE003586 Drosophila melan

seq_name: gb_pat2:AX083412

seq_documentation_block:
LOCUS       AX083412      681 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 104 from Patent WO0112660.
ACCESSION   AX083412
VERSION     AX083412.1  GI:13185249
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 104 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      179 a    186 c    183 g    133 t
ORIGIN


alignment_scores:
           Quality: 1115.00             Length:    205
             Ratio:  5.439               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AX083412   ..

 Align seg 1/1  to: AX083412  from: 1  to: 681

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 116

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 166

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 216

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 266

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 316

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 366

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 416

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 466
```

```
    134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    467 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 516

    151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    517 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 566

    167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    567 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 616

    184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    617 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 666

    201 GluIleAlaAlaCys 205
        ||||||||||||||||
    667 GAGATAGCTGCCTGC 681

seq_name: gb_pat2:AX083422

seq_documentation_block:
LOCUS       AX083422      897 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 114 from Patent WO0112660.
ACCESSION   AX083422
VERSION     AX083422.1  GI:13185262
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0112660-A 114 22-FEB-2001;
            SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             99. .782
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC33305.1"
                     /db_xref="GI:13185263"
                     /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
                     GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                     AEPRQRFWRHWLVTDIKGADLKKGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                     VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                     KNQAEIAAC"
BASE COUNT      226 a    253 c    234 g    184 t
ORIGIN


alignment_scores:
            Quality: 1115.00            Length:    205
              Ratio:  5.439              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AX083422   ..

 Align seg 1/1  to: AX083422  from: 1  to: 897

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 214

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 264

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 314

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 364

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 414

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 464

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 514

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     515 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 564

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 614

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 664

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     665 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 714

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     715 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 764

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     765 GAGATAGCTGCCTGC 779

seq_name: gb_pat1:AX060293

seq_documentation_block:
LOCUS       AX060293      932 bp    DNA             PAT       22-JAN-2001
DEFINITION  Sequence 1 from Patent WO0078802.
ACCESSION   AX060293
VERSION     AX060293.1  GI:12405782
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and
            Herrmann,J.L.
  TITLE     Secreted polypeptides and corresponding polynucleotides
  JOURNAL   Patent: WO 0078802-A 1 28-DEC-2000;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             113. .796
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC24866.1"
                     /db_xref="GI:12405783"
                     /translation="MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQ
```

```
                    GLEVFYPELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSR
                    AEPRQRFWRHWLVTDIKGADLKEGKIQGQELSAYQAPSPPAHSGFHRYQFFVYLQEGK
                    VISLLPKENKTRGSWKMDRFLNRFHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKH
                    KNQAEIAAC"
BASE COUNT      244 a    257 c    244 g    187 t
ORIGIN


alignment_scores:
           Quality: 1111.00            Length:    205
             Ratio:  5.420             Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x AX060293   ..

 Align seg 1/1  to: AX060293  from: 1  to: 932

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     179 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 228

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     229 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 278

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     279 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 328

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     329 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 378

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     379 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 428

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     429 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 478

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||:::||||||||||||||||||||||||||||||||||||||||||
     479 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 528

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     529 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 578

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     579 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 628

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     629 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 678

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     679 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 728

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     729 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 778

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     779 GAGATAGCTGCCTGC 793

seq_name: gb_sts2:G27363

seq_documentation_block:
```

```
LOCUS       G27363        352 bp    DNA             STS       28-JUN-1996
DEFINITION  human STS SHGC-31033, sequence tagged site.
ACCESSION   G27363
VERSION     G27363.1  GI:1396086
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: TTGGAAAAGAAGGGGTTCTG
            Primer B: AAGCACAAAAACCAGGCG
            STS size: 133
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from H83927
            -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="8"
     STS             4. .136
     primer_bind     4. .23
     primer_bind     complement(119. .136)
BASE COUNT       69 a     76 c    110 g     96 t      1 others
ORIGIN


alignment_scores:
           Quality:  444.00             Length:     83
             Ratio:  5.349                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x G27363/rev   ..

 Align seg 1/1  to reverse of: G27363  from: 1  to: 352

     123 SerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysVa 139
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 AGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCTTCAGGAAGGAAAAGT 303
```

```
    139 lIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetA 156
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 CATCTCTCTCCTTCCCAAGGAAAACAAAACTCGAGGCTCTTGGAAAATGG 253

    156 spArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGln 172
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 ACAGATTTCTGAACCGTTTCCACCTGGGCGAACCTGAAGCAAGCACCCAG 203

    173 PheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaProArgGl 189
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TTCATGACCCAGAACTACCAGGACTCACCAACCCTCCAGGCTCCCAGAGA 153

    189 uArgAlaSerGluProLysHisLysAsnGlnAlaGluIleAlaAlaCys 205
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 AAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCGGAGATAGCTGCCTGC 104

seq_name: gb_htg13:AC055854

seq_documentation_block:
LOCUS       AC055854   179814 bp    DNA             HTG       25-MAY-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-459E5 map 8, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC055854
VERSION     AC055854.3  GI:8084680
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179814)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-459E5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179814)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7770648.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9765
                Center clone name: 459_E_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167541 bases at least Q40
                Consensus quality: 173729 bases at least Q30
                Consensus quality: 175997 bases at least Q20
                Insert size: 175000; agarose-fp
                Insert size: 177314; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1730: contig of 1730 bp in length
            *     1731 1830: gap of      100 bp
            *     1831     3230: contig of 1400 bp in length
            *     3231 3330: gap of      100 bp
            *     3331     5675: contig of 2345 bp in length
            *     5676 5775: gap of      100 bp
            *     5776     7874: contig of 2099 bp in length
            *     7875 7974: gap of      100 bp
            *     7975    10477: contig of 2503 bp in length
            *    10478 10577: gap of      100 bp
            *    10578    12553: contig of 1976 bp in length
            *    12554 12653: gap of      100 bp
            *    12654    16386: contig of 3733 bp in length
            *    16387 16486: gap of      100 bp
            *    16487    19715: contig of 3229 bp in length
            *    19716 19815: gap of      100 bp
            *    19816    23166: contig of 3351 bp in length
            *    23167 23266: gap of      100 bp
            *    23267    26855: contig of 3589 bp in length
            *    26856 26955: gap of      100 bp
            *    26956    31690: contig of 4735 bp in length
            *    31691 31790: gap of      100 bp
            *    31791    38834: contig of 7044 bp in length
            *    38835 38934: gap of      100 bp
            *    38935    46670: contig of 7736 bp in length
            *    46671 46770: gap of      100 bp
            *    46771    52635: contig of 5865 bp in length
            *    52636 52735: gap of      100 bp
            *    52736    59853: contig of 7118 bp in length
            *    59854 59953: gap of      100 bp
            *    59954    66887: contig of 6934 bp in length
            *    66888 66987: gap of      100 bp
            *    66988    75106: contig of 8119 bp in length
            *    75107 75206: gap of      100 bp
            *    75207    84076: contig of 8870 bp in length
            *    84077 84176: gap of      100 bp
            *    84177    91487: contig of 7311 bp in length
            *    91488 91587: gap of      100 bp
            *    91588    98734: contig of 7147 bp in length
            *    98735 98834: gap of      100 bp
            *    98835   107787: contig of 8953 bp in length
            *   107788 107887: gap of      100 bp
            *   107888   118191: contig of 10304 bp in length
            *   118192 118291: gap of      100 bp
            *   118292   134239: contig of 15948 bp in length
            *   134240 134339: gap of      100 bp
            *   134340   148401: contig of 14062 bp in length
            *   148402 148501: gap of      100 bp
            *   148502   163397: contig of 14896 bp in length
            *   163398 163497: gap of      100 bp
            *   163498   179814: contig of 16317 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179814
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
```

```
                     /map="8"
                     /clone="RP11-459E5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1730
                     /note="assembly_fragment"
     misc_feature    1831. .3230
                     /note="assembly_fragment"
     misc_feature    3331. .5675
                     /note="assembly_fragment"
     misc_feature    5776. .7874
                     /note="assembly_fragment"
     misc_feature    7975. .10477
                     /note="assembly_fragment"
     misc_feature    10578. .12553
                     /note="assembly_fragment"
     misc_feature    12654. .16386
                     /note="assembly_fragment"
     misc_feature    16487. .19715
                     /note="assembly_fragment"
     misc_feature    19816. .23166
                     /note="assembly_fragment"
     misc_feature    23267. .26855
                     /note="assembly_fragment"
     misc_feature    26956. .31690
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    31791. .38834
                     /note="assembly_fragment"
     misc_feature    38935. .46670
                     /note="assembly_fragment"
     misc_feature    46771. .52635
                     /note="assembly_fragment"
     misc_feature    52736. .59853
                     /note="assembly_fragment"
     misc_feature    59954. .66887
                     /note="assembly_fragment"
     misc_feature    66988. .75106
                     /note="assembly_fragment"
     misc_feature    75207. .84076
                     /note="assembly_fragment"
     misc_feature    84177. .91487
                     /note="assembly_fragment"
     misc_feature    91588. .98734
                     /note="assembly_fragment"
     misc_feature    98835. .107787
                     /note="assembly_fragment"
     misc_feature    107888. .118191
                     /note="assembly_fragment"
     misc_feature    118292. .134239
                     /note="assembly_fragment"
     misc_feature    134340. .148401
                     /note="assembly_fragment"
     misc_feature    148502. .163397
                     /note="assembly_fragment"
     misc_feature    163498. .179814
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
BASE COUNT    43709 a  43721 c  45547 g  44332 t   2505 others
ORIGIN


alignment_scores:
           Quality:  262.50            Length:    137
             Ratio:  3.052               Gaps:      6
 Percent Similarity:  62.774   Percent Identity: 51.095

alignment_block:
 US-09-215-435-225_COPY_23_227 x AC055854/rev   ..

 Align seg 1/1  to reverse of: AC055854  from: 1  to: 179814
```

```
       74 AspProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHi 90
          :::|||:::||||||      :::   |||
    76358 GAGCCGGAAGCTCCCCCAGCCACTACGCCC..................... 76329

       90 sTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleG 107
                                                           :
    76328 .............................................C 76328

      107 lnGlyGlnGluLeuSer.AlaTyrGlnAlaProSer......ProPro.. 120
          ::|||:::::::::::: |||   :::||||||:::      ||||||
    76327 ATGGCCATCGCATCACCTGCTGCCTCGGCTCCCACGAGCCGGCCTCCCTC 76278

      121 ....AlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnG 136
              |||   ::::::|||:::   :::      |||
    76277 TCAGGCCTCTGCCAGCTTTCAAGTCCATGGCTCTTTCTCACTGCCTCCCT 76228

      136 luGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySer 152
            |||      :::   :::||||||                    ||||||
    76227 GTGGT......CTGTTTGTCCTTCCA.................GGCTCT 76202

      153 TrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAl 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    76201 TGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACCTGAAGC 76152

      169 aSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnA 186
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    76151 AAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCCTCCAGG 76102

      186 laProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaGluIle 202
          ||||||||   ||||||||||||||||||||   ||||||||||||||||
    76101 CTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.AACCAGGCGGAGATA 76053

      203 AlaAlaCys 205
          |||||||||
    76052 GCTGCCTGC 76044

seq_name: gb_ro2:RNPBP

seq_documentation_block:
LOCUS       RNPBP        1037 bp    mRNA            ROD       17-FEB-1997
DEFINITION  R.norvegicus mRNA for phosphatidylethanolamine-binding protein.
ACCESSION   X71873
VERSION     X71873.1  GI:510338
KEYWORDS    epididymal protein; pbp gene; phosphatidylethanolamine-binding
            protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   Perry,A.C., Hall,L., Bell,A.E. and Jones,R.
  TITLE     Sequence analysis of a mammalian phospholipid-binding protein from
            testis and epididymis and its distribution between spermatozoa and
            extracellular secretions
  JOURNAL   Biochem. J. 301 (Pt 1), 235-242 (1994)
  MEDLINE   94311839
REFERENCE   2  (bases 1 to 1037)
  AUTHORS   Hall,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1993) L. Hall, University of Bristol, Dept of
            Biochemistry, School of Medical Sciences, University Walk, Bristol
            BS8 1TD, UK
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /dev_stage="adult male"
                     /tissue_type="epididymis and liver"
                     /clone_lib="cDNA"
                     /clone="18 clones"
     5'UTR           1. .28
     gene            29. .592
```

```
                /gene="PBP"
     CDS        29. .592
                /gene="PBP"
                /codon_start=1
                /product="phosphatidylethanolamine-binding protein"
                /protein_id="CAA50708.1"
                /db_xref="GI:510339"
                /db_xref="SWISS-PROT:P31044"
                /translation="MAADISQWAGPLSLQEVDEPPQHALRVDYGGVTVDELGKVLTPT
                QVMNRPSSISWDGLDPGKLYTLVLTDPDAPSRKDPKFREWHHFLVVNMKGNDISSGTV
                LSEYVGSGPPKDTGLHRYVWLVYEQEQPLNCDEPILSNKSGDNRGKFKVESFRKKYHL
                GAPVAGTCFQAEWDDSVPKLHDQLAGK"
     3'UTR      590. .1037
     variation  766
                /note="in 6 clones"
                /replace="t"
BASE COUNT      244 a    299 c    290 g    204 t
ORIGIN


alignment_scores:
            Quality:  237.50            Length:    185
              Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225_COPY_23_227 x RNPBP   ..

 Align seg 1/1  to: RNPBP  from: 1  to: 1037

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         ||||||||||::::::        ::::::|||                ||
     134 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 174

      46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
         |   :::      ||||||                 |||
     175 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 203

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
           |||      |||   ||||||:::::|||||||||||||||||||
     204 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 253

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     254 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 303

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||:::::::::|||   :::             ||||||
     304 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 341

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
           |||   :::   :::|||||||      :::|||:::|||||||||
     342 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 391

     130 PhePheValTyrLeuGlnGluGlyLysValIleSer.............Le 142
         ::::::||||||   ||||||      :::                 ||
     392 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 441

     142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
         |      |||::::::   :::||||||:::::|||:::::   |||
     442 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 491

     159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     492 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 541

     176 GlnAsnTyrGlnAspSer.............................. 181
         :::   ::::::||||||
     542 GAG...TGGGATGACTCTGTGCCCAAGCTGCACGATCAGCTGGCTGGGAA 588

     182 ................................ProThrLeuGlnAlaP 187
                                        ||||||:::|||:::
     589 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 638
```

```
     187 roArg 188
            |||
     639 GTCGT 643

seq_name: gb_pat2:E05646

seq_documentation_block:
LOCUS       E05646       1047 bp    RNA             PAT       29-SEP-1997
DEFINITION  cDNA encoding rat hippocampal cholinergic neurostimulating
            peptide,HCNP.
ACCESSION   E05646
VERSION     E05646.1  GI:2173833
KEYWORDS    JP 1993268966-A/1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   Tofuji,N., Tojo,S., Kojima,S., Ueki,Y., Nishihara,N., Fukushima,N.,
            Irie,T., Ono,K., Agui,H. and Kojika,Y.
  TITLE     NEUROTROPHIC FACTOR-RELATED POLYPEPTIDE AND GENE USEFUL THEREFOR
  JOURNAL   Patent: JP 1993268966-A 1 19-OCT-1993;
            SUMITOMO PHARMACEUT CO LTD, KOJIKA YUKIO, YAMAMOTO MASAHIKO
COMMENT     OS   Rattus norvegicus (rat)
            PN   JP 1993268966-A/1
            PD   19-OCT-1993
            PF   25-APR-1992 JP 1992131703
            PR   27-APR-1991 JP 91P   124688
            PI   TOFUJI NAOKI, TOJO SHINICHIRO, KOJIMA SHINICHI, UEKI YASUYUKI,
            PI   NISHIHARA NORIO, FUKUSHIMA NOBUYUKI, IRIE TSUNEMASA, PI   ONO
            KEIICHI,
            PI   AGUI HIDEO, KOJIKA YUKIO
            PC   C12N15/12,A61K37/02,A61K37/02,C07K7/06,C07K13/00,C12N1/21, PC
             C12N5/10,
            PC   (C12N1/21,C12R1:19),(C12N5/10,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: tissue_type=hippocampus of brain;
            CC   *source: clone=AO10-12;
            CC   Feature is identified by similarity;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           <1. .25
            FT   CDS             26. .589
            FT                   /product='hippocampal cholinergic  FT
                    neurostimulating peptide'
            FT   3'UTR           590. .1047
            FT   polyA_signal    987. .992
            FT   polyA_site      1008. .1047
            FT   mutation        replace(427. .429,'AAG')
            FT                   /note='hippocampal cholinergic  FT
                  neurostimulating peptide
            FT                   derived from clone AO1-1'.
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
BASE COUNT      257 a    297 c    289 g    204 t
ORIGIN


alignment_scores:
            Quality:  237.50            Length:    185
              Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225_COPY_23_227 x E05646   ..
```

```
Align seq 1/1  to: E05646  from: 1  to: 1047

     30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
        |||||||||::::::       ::::::|||                ||
    131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

     46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
        |   :::      ||||||                  |||
    172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

     63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
           |||      |||   ||||||::::::||||||||||||||||||
    201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

     80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
        |||   :::|||:::   |||   |||:::|||:::||||||:::::::
    251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

     96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
        :||||||   |||:::::::::|||   :::             ||||||
    301 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 338

    113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
          |||   :::   :::||||||      :::|||:::|||||||||
    339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

    130 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 142
        ::::::||||||   ||||||      :::                  ||
    389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

    142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
        |      |||::::::   :::|||||||:::::|||:::::   |||
    439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

    159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
          :::::::::|||||||||   |||   |||:::|||   |||::::::
    489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

    176 GlnAsnTyrGlnAspSer.............................. 181
        :::   ::::::||||||
    539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585

    182 ..............................ProThrLeuGlnAlaP 187
                                      ||||||:::|||:::
    586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

    187 roArg 188
          |||
    636 GTCGT 640

seq_name: gb_ro2:RNPEABP

seq_documentation_block:
LOCUS       RNPEABP      1075 bp    mRNA            ROD       25-AUG-1995
DEFINITION  R.norvegicus phosphatidylethanolamine binding protein mRNA.
ACCESSION   X75253
VERSION     X75253.1  GI:406291
KEYWORDS    phosphatidylethanolamine-binding protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1075)
  AUTHORS   Tohdoh,N., Tojo,S., Agui,H. and Ojika,K.
  TITLE     Sequence homology of rat and human HCNP precursor proteins, bovine
            phosphatidylethanolamine-binding protein and rat 23-kDa protein
            associated with the opioid-binding protein
  JOURNAL   Brain Res. Mol. Brain Res. 30 (2), 381-384 (1995)
  MEDLINE   95364631
REFERENCE   2  (bases 1 to 1075)
  AUTHORS   Tohdoh,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1993) N. Tohdoh, Sumitomo Pharmaceuticals Co.,
            Ltd., 1-98, Kasugade-Naka 3-chome, Konohanaku, Osaka, 554, JAPAN
FEATURES             Location/Qualifiers
     source          1. .1075
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /germline
                     /tissue_type="hippocampus"
                     /clone_lib="lambda gt11"
                     /clone="pA01-1, pA010-1, pA62"
     mRNA            1. .1075
     CDS             54. .617
                     /codon_start=1
                     /product="phosphatidylethanolamine binding protein"
                     /protein_id="CAA53032.1"
                     /db_xref="GI:406292"
                     /db_xref="SWISS-PROT:P31044"
                     /translation="MAADISQWAGPLSLQEVDEPPQHALRVDYGGVTVDELGKVLTPT
                     QVMNRPSSISWDGLDPGKLYTLVLTDPDAPSRKDPKFREWHHFLVVNMKGNDISSGTV
                     LSEYVGSGPPKDTGLHRYVWLVYEQEQPLNCDEPILSNKSGDNRGKFKVESFRKKYHL
                     GAPVAGTCFQAEWDDSVPKLHDQLAGK"
     polyA_signal    1015. .1020
     polyA_signal    1036. .1075
BASE COUNT      261 a    305 c    298 g    211 t
ORIGIN


alignment_scores:
           Quality:  237.50            Length:    185
             Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225_COPY_23_227 x RNPEABP   ..

 Align seg 1/1  to: RNPEABP  from: 1  to: 1075

     30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
        |||||||||::::::       ::::::|||                ||
    159 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 199

     46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
        |   :::      ||||||                  |||
    200 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 228

     63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
           |||      |||   ||||||::::::||||||||||||||||||
    229 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 278

     80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
        |||   :::|||:::   |||   |||:::|||:::||||||:::::::
    279 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 328

     96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
        :||||||   |||:::::::::|||   :::             ||||||
    329 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 366

    113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
          |||   :::   :::||||||      :::|||:::|||||||||
    367 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 416

    130 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 142
        ::::::||||||   ||||||      :::                  ||
    417 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 466

    142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
        |      |||::::::   :::|||||||:::::|||:::::   |||
    467 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 516

    159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
          :::::::::|||||||||   |||   |||:::|||   |||::::::
    517 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 566
```

```
    176 GlnAsnTyrGlnAspSer.............................. 181
        :::   ::::::||||||
    567 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 613

    182 ..............................ProThrLeuGlnAlaP 187
                                      ||||||:::|||:::
    614 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 663

    187 roArg 188
          |||
    664 GTCGT 668
```

seq_name: gb_htg12:AC037441

seq_documentation_block:

LOCUS AC037441 162795 bp DNA HTG 03-MAR-2001
DEFINITION Homo sapiens clone RP11-87E22, WORKING DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AC037441
VERSION AC037441.3 GI:13184260
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162795)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-87E22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162795)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 2, 2001 this sequence version replaced gi:8705171.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L9203
Center clone name: 87_E_22
---------- Summary Statistics
Sequencing vector: M13; M77815; 81% of reads
Sequencing vector: Plasmid; n/a; 19% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157025 bases at least Q40
Consensus quality: 160134 bases at least Q30
Consensus quality: 161093 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 161795; sum-of-contigs
Quality coverage: 5.4 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 ba.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15243: contig of 15243 bp in length
* 15244 15343: gap of 100 bp
* 15344 18897: contig of 3554 bp in length
* 18898 18997: gap of 100 bp
* 18998 25211: contig of 6214 bp in length
* 25212 25311: gap of 100 bp
* 25312 34838: contig of 9527 bp in length
* 34839 34938: gap of 100 bp
* 34939 45351: contig of 10413 bp in length
* 45352 45451: gap of 100 bp
* 45452 57259: contig of 11808 bp in length
* 57260 57359: gap of 100 bp
* 57360 67990: contig of 10631 bp in length
* 67991 68090: gap of 100 bp
* 68091 83399: contig of 15309 bp in length
* 83400 83499: gap of 100 bp
* 83500 105395: contig of 21896 bp in length
* 105396 105495: gap of 100 bp
* 105496 129959: contig of 24464 bp in length
* 129960 130059: gap of 100 bp
* 130060 162795: contig of 32736 bp in length.

FEATURES Location/Qualifiers

source	1. .162795 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="RP11-87E22" /clone_lib="RPCI-11 Human Male BAC"
misc_feature	1. .15243 /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature	15344. .18897 /note="assembly_fragment"
misc_feature	18998. .25211 /note="assembly_fragment"
misc_feature	25312. .34838 /note="assembly_fragment"
misc_feature	34939. .45351 /note="assembly_fragment"
misc_feature	45452. .57259 /note="assembly_fragment"
misc_feature	57360. .67990 /note="assembly_fragment"
misc_feature	68091. .83399 /note="assembly_fragment"
misc_feature	83500. .105395 /note="assembly_fragment"
misc_feature	105496. .129959 /note="assembly_fragment"
misc_feature	130060. .162795 /note="assembly_fragment clone_end:T7 vector_side:right"

BASE COUNT 42087 a 39347 c 38304 g 42057 t 1000 others
ORIGIN

alignment_scores:
Quality: 236.00 Length: 43

```
            Ratio:  5.488                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.674

alignment_block:
 US-09-215-435-225_COPY_23_227 x AC037441/rev   ..

 Align seg 1/1  to reverse of: AC037441  from: 1  to: 162795

      22 GlnGlyLeuGluValPheTyrProGluLeuGlyAsnIleGlyCysLysVa 38
         :::||||||||||||||||||||||||||||||||||||||||||||||||
  151900 AGGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACATTGGCTGCAAGGT 151851

      38 lValProAspCysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluP 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  151850 TGTTCCTGATTGTAACAACTACAGACAGAAGATCACCTCCTGGATGGAGC 151801

      55 roIleValLysPheProGlyAlaValAsp 64
         |||||||||||||||||||||||||||||
  151800 CGATAGTCAAGTTCCCGGGGGCCGTGGAC 151772

seq_name: gb_htg17:AC084081

seq_documentation_block:
LOCUS       AC084081   169645 bp    DNA             HTG       11-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-89M8 map 8, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC084081
VERSION     AC084081.3  GI:13273380
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-89M8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2001 this sequence version replaced gi:12583824.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11307
                Center clone name: 89_M_8
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162624 bases at least Q40
                Consensus quality: 165742 bases at least Q30
                Consensus quality: 167046 bases at least Q20
                Insert size: 169000; agarose-fp
                Insert size: 168045; sum-of-contigs
                Quality coverage: 5.9 in Q20 bases; agarose-fp
                Quality coverage: 5.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      335: contig of 335 bp in length
            *      336 435: gap of      100 bp
            *      436     2289: contig of 1854 bp in length
            *     2290 2389: gap of      100 bp
            *     2390     4997: contig of 2608 bp in length
            *     4998 5097: gap of      100 bp
            *     5098    64677: contig of 59580 bp in length
            *    64678 64777: gap of      100 bp
            *    64778    68151: contig of 3374 bp in length
            *    68152 68251: gap of      100 bp
            *    68252    72655: contig of 4404 bp in length
            *    72656 72755: gap of      100 bp
            *    72756    78893: contig of 6138 bp in length
            *    78894 78993: gap of      100 bp
            *    78994    84664: contig of 5671 bp in length
            *    84665 84764: gap of      100 bp
            *    84765    90853: contig of 6089 bp in length
            *    90854 90953: gap of      100 bp
            *    90954    98425: contig of 7472 bp in length
            *    98426 98525: gap of      100 bp
            *    98526   106371: contig of 7846 bp in length
            *   106372 106471: gap of      100 bp
            *   106472   115673: contig of 9202 bp in length
            *   115674 115773: gap of      100 bp
            *   115774   127138: contig of 11365 bp in length
            *   127139 127238: gap of      100 bp
            *   127239   138436: contig of 11198 bp in length
            *   138437 138536: gap of      100 bp
            *   138537   149766: contig of 11230 bp in length
            *   149767 149866: gap of      100 bp
            *   149867   167340: contig of 17474 bp in length
            *   167341 167440: gap of      100 bp
            *   167441   169645: contig of 2205 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-89M8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .335
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    436. .2289
                     /note="assembly_fragment"
     misc_feature    2390. .4997
                     /note="assembly_fragment"
     misc_feature    5098. .64677
                     /note="assembly_fragment"
     misc_feature    64778. .68151
                     /note="assembly_fragment"
```

```
     misc_feature    68252. .72655
                     /note="assembly_fragment"
     misc_feature    72756. .78893
                     /note="assembly_fragment"
     misc_feature    78994. .84664
                     /note="assembly_fragment"
     misc_feature    84765. .90853
                     /note="assembly_fragment"
     misc_feature    90954. .98425
                     /note="assembly_fragment"
     misc_feature    98526. .106371
                     /note="assembly_fragment"
     misc_feature    106472. .115673
                     /note="assembly_fragment"
     misc_feature    115774. .127138
                     /note="assembly_fragment"
     misc_feature    127239. .138436
                     /note="assembly_fragment"
     misc_feature    138537. .149766
                     /note="assembly_fragment"
     misc_feature    149867. .167340
                     /note="assembly_fragment"
     misc_feature    167441. .169645
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44253 a  41074 c  39720 g  42960 t   1638 others
ORIGIN


alignment_scores:
            Quality:  236.00            Length:     43
              Ratio:  5.488               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.674

alignment_block:
 US-09-215-435-225_COPY_23_227 x AC084081/rev   ..

 Align seg 1/1  to reverse of: AC084081  from: 1  to: 169645

      22 GlnGlyLeuGluValPheTyrProGluLeuGlyAsnIleGlyCysLysVa 38
         :::||||||||||||||||||||||||||||||||||||||||||||||||
   41905 AGGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACATTGGCTGCAAGGT 41856

      38 lValProAspCysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluP 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   41855 TGTTCCTGATTGTAACAACTACAGACAGAAGATCACCTCCTGGATGGAGC 41806

      55 roIleValLysPheProGlyAlaValAsp 64
         |||||||||||||||||||||||||||||
   41805 CGATAGTCAAGTTCCCGGGGGCCGTGGAC 41777

seq_name: gb_in4:OVD1PROTN

seq_documentation_block:
LOCUS       OVD1PROTN     611 bp    mRNA            INV       05-NOV-1996
DEFINITION  O.volvulus mRNA for d1 protein.
ACCESSION   X87991
VERSION     X87991.1  GI:1143526
KEYWORDS    D1 gene; D1 protein.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Erttmann,K.D. and Gallin,M.Y.
  TITLE     Onchocerca volvulus: identification of cDNAs encoding a putative
            phosphatidyl-ethanolamine-binding protein and a putative partially
            processed mRNA precursor
  JOURNAL   Gene 174 (2), 203-207 (1996)
  MEDLINE   97045813
REFERENCE   2  (bases 1 to 611)
  AUTHORS   Erttmann,K.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1995) K.D. Erttmann, Bernhard Nocht Inst for
            Tropical Med., Bernhard-Nocht-Str. 74, 20359 Hamburg, FRG
COMMENT     Related sequences M27807, U05980, U05981, U05982, U05984, U05985,
            U05243, U05244.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Onchocerca volvulus"
                     /db_xref="taxon:6282"
                     /dev_stage="adult"
                     /clone_lib="lambda ZAP"
                     /cell_line="E.Coli"
     gene            1. .459
                     /gene="d1"
     CDS             <1. .459
                     /gene="d1"
                     /codon_start=1
                     /product="D1 protein"
                     /protein_id="CAA61244.1"
                     /db_xref="GI:1143527"
                     /db_xref="SWISS-PROT:P54186"
                     /translation="VVSTSPTKLVNVSYNNLTVNLGNELTPTQVKNQPTKVSWDAEPG
                     ALYTLVMTDPDAPSRKNPVFREWHHWLIINISGQNVSSGTVLSDYIGSGPPKGTGLHR
                     YVFLVYKQPGSITDTQHGGNRPNFKVMDFANKHHLGNPVAGNFFQAKHED"
BASE COUNT      196 a    114 c    123 g    178 t
ORIGIN


alignment_scores:
            Quality:  234.50            Length:    152
              Ratio:  2.393               Gaps:      6
 Percent Similarity:  64.474   Percent Identity: 36.842

alignment_block:
 US-09-215-435-225_COPY_23_227 x OVD1PROTN   ..

 Align seg 1/1  to: OVD1PROTN  from: 1  to: 611

      33 AsnIleGlyCysLysValValPro...AspCysAsnAsnTyrArgGlnLy 48
         |||:::|||  :::::::::|||   :::   :::|||         ||
      58 AATCTGGGCAATGAACTTACGCCGACGCAGGTAAAGAATCAGCCGACAAA 107

      48 sIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAspG 65
         |:::  ||||||                                 |||     |
     108 AGTA...TCATGG.....................GATGCGGAACCTG 130

      65 lyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAla 81
         |||||  |||  |||||||||:::|||||||||||||||||||||||||
     131 GAGCCTTATATACGCTCGTTATGACTGATCCGGACGCACCATCTCGAAAA 180

      82 GluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGl 98
         :::|||      |||  |||:::||||||||:::  :::|||:::||
     181 AACCCCGTATTCAGAGAGTGGCACCATTGGTTGATAATTAATATTTCTGG 230

      98 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrG 115
         |  :::::::::::::|||  :::           ||||||  |||
     231 ACAAAATGTTAGCAGTGGCACAGTG...........TTATCTGATTATA 268

     115 lnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePhe 131
         :::  :::|||||      :::|||:::||||||||  |||:::
     269 TTGGATCAGGTCCACCAAAAGGCACAGGACTTCATCGTTATGTATTCTTG 318

     132 ValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLy 148
         ||||||  |||  |||::::::  :::         :::::::::
     319 GTTTATAAACAACCTGGAAGTATCACGGAT......ACTCAACATGGCGG 362

     148 sThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuG 165
         :::|||  :::::::|||:::      |||  |||:::  |||||||
     363 AAATCGCCCAAATTTCAAAGTTATGGATTTTGCAAACAAACATCACTTGG 412

     165 lyGluProGluAlaSerThrGlnPheMetThrGlnAsn.........Tyr 178
         ||:::|||  |||::::::  |||:::::::::::::         |||
     413 GAAATCCAGTTGCCGGGAACTTCTTCCAGGCTAAACATGAAGACTAATAT 462
```

```
    179 GlnAsp 180
        :::|||
    463 GAAGAC 468

seq_name: gb_in3:CELF40A3

seq_documentation_block:
LOCUS       CELF40A3    41155 bp    DNA             INV       02-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid F40A3.
ACCESSION   AF016423
VERSION     AF016423.1  GI:2291195
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41155)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 41155)
  AUTHORS   Geisel,C., Bradshaw,H. and Keppler,D.
  TITLE     The sequence of C. elegans cosmid F40A3
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 41155)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1997)
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone


                        NEIGHBORING COSMID INFORMATION:


            The 5' cosmid is C37C3, 200 bp overlap;3' lies in a gap.  Actual
            start of this cosmid is at base position 197 of CELF40A3; actual
            end is at 41155 of CELF40A3

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES             Location/Qualifiers
     source          1. .41155
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="F40A3"
     gene            complement(156. .1605)
                     /gene="F40A3.5"
     CDS             complement(join(156. .216,268. .662,714. .787,864. .981,
                     1094. .1281,1357. .1431,1476. .1605))
                     /gene="F40A3.5"
                     /note="contains similarity to a protein kinase domain"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65324.1"
                     /db_xref="GI:2291201"
                     /translation="MELDHTPPPSVLNDNCSASYMTPYATVIAMSGLYLLAIFYFCKK
                     SKKMCQPMSDSIYPYQKRLTQLERELFKDVRAHSVFFAELKNYLIDEESIEVDDFQIG
                     QTADGFIFRGGVFPKTRNRFNAKVTTAVKISFPIVSKSISLLEDALRLSKLDHPNLIR
                     LLAVSQLSFTVFRPMIALEWLPGGTLADYFQFKVREKDDSERSPIQLKDMLSILYQVS
                     QALKYIHSQLDEFGQELTHGRIFTRNVLVTEPDLRKCEVKLGDFGDAPMGLEYSTPII
                     AYMPPEILCCAERIPPHRPENDVWMFGVFIWECLTLGAQPHFRKSVEEIKKSFRLPDR
                     GLSCPPTCPLDV"
     gene            6740. .7594
                     /gene="F40A3.4"
     CDS             join(6740. .6827,7331. .7394,7450. .7594)
                     /gene="F40A3.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65323.1"
                     /db_xref="GI:2291200"
                     /translation="MSRMDSYSSSIREEQKAILNGLLGILFADGIIYDNYDTTRKIHQ
                     TVDGIYRMVVALLLIVIALIIIAVFLICFFHCRRANNEERSPLHSNRKSHPIEI"
     gene            8506. .9283
                     /gene="F40A3.3"
     CDS             join(8506. .8600,8667. .8781,8828. .9283)
                     /gene="F40A3.3"
                     /note="Similar to phosphatidylethanolamine binding
                     protein; coded for by C. elegans cDNA yk106a12.5; coded
                     for by C. elegans cDNA yk106a12.3"
                     /codon_start=1
                     /protein_id="AAB65322.1"
                     /db_xref="GI:2291199"
                     /translation="MVVLVTRSLLPALFFASRAPFAAATTSARFQRGLATMAAEAFTK
                     HEVIPDVLASNPPSKVVSVKFNSGVEANLGNVLTPTQVKDTPEVKWDAEPGALYTLIK
                     TDPDAPSRKEPTYREWHHWLVVNIPGNDIAKGDTLSEYIGAGPPPKTGLHRYVYLIYK
                     QSGRIEDAEHGRLTNTSGDKRGGWKAADFVAKHKLGAPVFGNLFQAEYDDYVPILNKQ
                     LGA"
     gene            9824. .12167
                     /gene="F40A3.2"
     CDS             join(9824. .9917,10735. .11058,11837. .12034,12091. .12167)
                     /gene="F40A3.2"
                     /note="contains similarity to EGF-like domains"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65320.1"
                     /db_xref="GI:2291197"
                     /translation="MLRRILLSAILIVLVACVLRKKKPCINGTPEGDRCYCIEGWTGT
                     LCHRKMYCSGFERLQNGSCIECELGWAGTDCDIIDCHGNGMPNYDLTECTCTVPYSGK
                     YCEIADTKDIYKWYNNFTSSIGPIGILTIIPLALIYMCCEHFAKKRQLRRVEQHLNGV
                     YITTGRKQVDKMLVKGLLEASDDEKSVKSVKIAADSGKEDKDKEEEEVIKMRNSATTS
                     EKKALLADPADM"
     gene            complement(23146. .24752)
                     /gene="F40A3.6"
     CDS             complement(join(23146. .23311,23512. .23597,23644. .23841,
                     23885. .24052,24594. .24752))
                     /gene="F40A3.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65325.1"
                     /db_xref="GI:2291202"
                     /translation="MSEYLGWHNRKPAAGSRGNSSFDRDIVPGGAHTKSYESYVKNIF
```

```
                HPTNRLRAKDKDQLVRDHGLSATSMLRAEREFGRSALDDKYDREHFFEKMPLESLYAE
                LGLKNPAASNYRSFQVYEPKSSSEYFAASYEAERERRKKDLERELASYDKEDYEDFTG
                EPKKNLYSSISSYRPPLSSYKSREDEEYEKNNRKTLIDYKDFDSSPLSPYLYQSHPYV
                RSQDSRIIGSNFVQLTSRPKDKFISKIEETLAMVRDMPRY"
    gene        24968. .26143
                /gene="F40A3.1"
    CDS         join(24968. .25107,25199. .25454,25498. .25634,26041. .26143)
                /gene="F40A3.1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB65321.1"
                /db_xref="GI:2291198"
                /translation="MSEIVRLPEETQETEMPLHHANRAFGARKQDAGWMAKNEGKTFR
                VVIKNIEIFQINDWELIGSRTYSWKGNEVLRCKHWDKVAIRWHLSTCTEHGVCDGGTN
                VDYATRAQNWTEMPVWRNAFTYVQIPGQLQKTKFLRLGLTGYDSRKTQLEAFSPLKID
                VDWRWQVTRDGCNNFVVRMSAMSMINEKMRPKLERFCRLFDKDHFVDCTNL"
    gene        complement(26520. .30751)
                /gene="F40A3.7"
    CDS         complement(join(26520. .26710,27914. .28026,28197. .28358,
                28415. .28609,29618. .29761,29814. .29935,30397. .30616,
                30669. .30751))
                /gene="F40A3.7"
                /note="conatains similarity to seven-transmembrane
                receptors"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB65319.1"
                /db_xref="GI:2291196"
                /translation="MSANFSHLCLTDEQQRMSGTNLEYNLQRYFYPALAIFGILGNVL
                NLTVLLNRSMRSRANSFLAVLAFADIIFLFLLFPNILANYSFFTFNWYFRWFYLHSKV
                HLISLANWCSSVAHWCVIAVCGDRLLGIQNPLYARATWRWWKLPLVTTIIVFTCGLLT
                CYQHFEYFCLVRSYCRETQLYSRCLPVNAEKWFGHRPNPFSQRYQSFIAICKLAHIFL
                MIILPIILLLFLNLTLLWALRKRQKHLSIGKDFNADRRQNDVHMQKTEHRVTLTVTFI
                VTMFTLTNGPSALVHLVMYATHEELYDLTMISSTLVICGKASNFILFCLGSKHFRLRL
                LKLTRKKINRKIDSITGSLVHTTKLSTASSRRTSMPIQEKKRSSSCAPLVAKNRALST
                VETIYQPLLNGNRKGDF"
BASE COUNT    13599 a   7323 c   6798 g  13435 t
ORIGIN


alignment_scores:
            Quality:  230.00            Length:    159
              Ratio:  2.447               Gaps:      7
 Percent Similarity:  59.119   Percent Identity: 37.107

alignment_block:
 US-09-215-435-225_COPY_23_227 x CELF40A3   ..

 Align seg 1/1  to: CELF40A3  from: 1  to: 41155

     27 PheTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAs 43
        |||      :::||||||||||:::         ::::::|||
   8822 TTTCAGGCTAATCTTGGAAACGTT.........CTCACTCCAACC..... 8857

     43 nAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPheP 60
           :::   |||   |||            |||   |||||||:::
   8858 .......CAAGTAAAGGATACT............CCAGAAGTAAAATGG. 8887

     60 roGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAsp 76
         |||      ||||||   |||   |||:::   :::||||||||||
   8888 ..GATGCTGAGCCAGGAGCACTCTACACTCTCATCAAGACGGATCCAGAT 8935

     77 AlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuVa 93
        ||||||||||||   ||||||      |||   |||:::||||||||||||
   8936 GCTCCATCTCGCAAAGAGCCAACCTACCGAGAGTGGCATCACTGGCTCGT 8985

     93 lThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnG 110
        |:::::|||   |||   |||:::  |||                |||:::
   8986 CGTCAACATTCCAGGAAACGACATTGCCAAA............GGAGACA 9023

    110 luLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHis 126
          ||||||   |||   :::   :::|||||||      :::|||:::|||
   9024 CTCTTTCCGAGTACATCGGCGCCGGACCACCACCAAAAACTGGACTCCAT 9073
```

```
    127 ArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer..... 141
        ||||||   ::::::::::|||   |||:::|||::::::   :::
   9074 CGTTACGTTTACCTGATCTACAAGCAATCTGGACGCATCGAAGATGCCGA 9123

    142 .........LeuLeuProLysGluAsnLysThrArgGlySerTrpLysM 155
                |||         ::::::      ||||||:::|||||||
   9124 GCACGGACGTCTAACCAACACTTCCGGAGACAAAAGAGGAGGATGGAAGG 9173

    155 etAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThr 171
           |||:::   :::      ||||||   |||      ::::::
   9174 CTGCCGATTTCGTCGCAAAGCACAAGCTCGGTGCTCCAGTCTTCGGAAAT 9223

    172 GlnPheMetThrGlnAsnTyrGlnAsp 180
        |||::::::::::   |||:::|||
   9224 CTGTTCCAGGCCGAG...TACGACGAC 9247

seq_name: gb_htg1:AC006915

seq_documentation_block:
LOCUS       AC006915   134095 bp    DNA              HTG       24-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y97E10y, *** SEQUENCING IN PROGRESS
            ***, 1 unordered pieces.
ACCESSION   AC006915
VERSION     AC006915.1  GI:4263433
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 134095)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134095)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   134095: contig of 134095 bp in length.
FEATURES            Location/Qualifiers
     source         1. .134095
                    /organism="Caenorhabditis elegans"
                    /db_xref="taxon:6239"
                    /clone="Y97E10y"
BASE COUNT    43230 a  24499 c  23702 g  42664 t
ORIGIN


alignment_scores:
            Quality:  230.00            Length:    159
              Ratio:  2.447               Gaps:      7
 Percent Similarity:  59.119   Percent Identity: 37.107

alignment_block:
 US-09-215-435-225_COPY_23_227 x AC006915   ..

 Align seg 1/1  to: AC006915  from: 1  to: 134095

     27 PheTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAs 43
        |||      :::||||||||||:::         ::::::|||
   99439 TTTCAGGCTAATCTTGGAAACGTT.........CTCACTCCAACC..... 99474

     43 nAsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPheP 60
```

```
            :::   |||   |||           |||   ||||||:::
 99475 .......CAAGTAAAGGATACT...........CCAGAAGTAAAATGG. 99504

    60 roGlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAsp 76
         |||      ||||||   |||   |||:::   :::||||||||||
 99505 ..GATGCTGAGCCAGGAGCACTCTACACTCTCATCAAGACGGATCCAGAT 99552

    77 AlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuVa 93
       ||||||||||||   ||||||      |||   |||:::|||||||||||
 99553 GCTCCATCTCGCAAAGAGCCAACCTACCGAGAGTGGCATCACTGGCTCGT 99602

    93 lThrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnG 110
       |:::::|||   |||   |||:::   |||             |||:::
 99603 CGTCAACATTCCAGGAAACGACATTGCCAAA...........GGAGACA 99640

   110 luLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHis 126
         ||||||   |||   :::   :::|||||||      :::|||:::|||
 99641 CTCTTTCCGAGTACATCGGCGCCGGACCACCACCAAAAACTGGACTCCAT 99690

   127 ArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSer..... 141
       ||||||   ::::::::::|||   |||:::|||::::::   :::
 99691 CGTTACGTTTACCTGATCTACAAGCAATCTGGACGCATCGAAGATGCCGA 99740

   142 ..........LeuLeuProLysGluAsnLysThrArgGlySerTrpLysM 155
                  |||         ::::::      ||||||:::||||||
 99741 GCACGGACGTCTAACCAACACTTCCGGAGACAAAAGAGGAGGATGGAAGG 99790

   155 etAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThr 171
             |||:::   :::      ||||||   |||      ::::::
 99791 CTGCCGATTTCGTCGCAAAGCACAAGCTCGGTGCTCCAGTCTTCGGAAAT 99840

   172 GlnPheMetThrGlnAsnTyrGlnAsp 180
         |||:::::::::   |||:::|||
 99841 CTGTTCCAGGCCGAG...TACGACGAC 99864
```

seq_name: gb_htg4:AC012978

```
seq_documentation_block:
LOCUS       AC012978    63551 bp    DNA             HTG       03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC012978
VERSION     AC012978.1  GI:6223345
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 63551)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212016 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .63551
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    18144 a  13546 c  14099 g  17762 t
ORIGIN


alignment_scores:
            Quality:  229.00            Length:    131
              Ratio:  2.516               Gaps:      5
 Percent Similarity:  69.466   Percent Identity: 40.458
```

```
alignment_block:
 US-09-215-435-225_COPY_23_227 x AC012978/rev   ..

 Align seg 1/1  to reverse of: AC012978  from: 1  to: 63551

     54 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 70
        ||||||   ||||||:::      |||      :::      |||   ||
  23000 GAGCCGTGCGTGAAATGGGAG...GCGGACGCCAATAAGCTGTACACCCT 22954

     70 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 87
        |   |||:::||||||||||||||||||||||   :::|||:::   |||
  22953 CTGCATGACCGATCCGGATGCGCCCAGTCGCAAGGATCCCAAGTTTAGGG 22904

     87 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 103
        |||:::||||||||||||   :::|||   |||:::|||:::   |||
  22903 AGTGGCACCATTGGCTGGTGGGCAACATACCCGGTGGAGATGTCGCCAAG 22854

    104 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 120
        |||::::::            ||||||||||||   :::   :::||||||
  22853 GGCGAGGTT...........CTCTCCGCCTACGTGGGATCCGGGCCTCC 22816

    120 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 137
        |      :::|||:::|||||||||   |||::::::|||   |||:::
  22815 ACCAGACACCGGACTCCATCGTTACGTTTTCCTGATCTACGAGCAGCGGT 22766

    137 lyLysValIle.........SerLeuLeuProLysGluAsn...LysThr 149
        |||:::          :::   ||||||:::::::::
  22765 GCAAGCTCACATTCGACGAGAAGCGACTGCCCAATAACAGCGGAGATGGA 22716

    150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
        ||||||:::::|||:::   :::|||   :::::::::   ||||||::
  22715 CGCGGTGGCTTCAAAATCGCCGAGTTCGCCAAGAAGTACGCCCTCGGCAA 22666

    166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAsp 180
        :|||   |||::::::   ::::::::::::   |||:::|||
  22665 TCCCATTGCCGGTAACCTCTACCAGGCTGAG...TACGACGAC 22626

seq_name: gb_in1:AC005891

seq_documentation_block:
LOCUS       AC005891   136933 bp    DNA             INV       30-OCT-1998
DEFINITION  Drosophila melanogaster, chromosome 2L, region 33F3-34A2, P1 clones
            DS08479 and DS07153, complete sequence.
ACCESSION   AC005891 AC005108 AC005109 AC005110
VERSION     AC005891.1  GI:3810575
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 136933)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2L, region 33F3-34A2
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 136933)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
```

```
            Zieran,L.L. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library locations: 89-31, 75-49.
FEATURES             Location/Qualifiers
     source          1. .136933
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="33F3-34A2"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
                     /clone="Pls DS08479 (D303) and DS07153 (D305)"
                     /note="DS08479 (D303) and DS07153 (D305) were completed as
                     a project.  DS08479 extends from P1 end at bp 1 to P1 end
                     at bp 83797.  DS07153  extends from p1 end at bp 51,253 to
                     P1 end at bp 136,933."
BASE COUNT    37380 a  28396 c  29033 g  42124 t
ORIGIN


alignment_scores:
            Quality:  229.00            Length:    131
              Ratio:  2.516               Gaps:      5
 Percent Similarity:  69.466   Percent Identity: 40.458

alignment_block:
 US-09-215-435-225_COPY_23_227 x AC005891   ..

 Align seg 1/1  to: AC005891  from: 1  to: 136933

      54 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 70
         ||||||   ||||||:::      |||      :::      |||   ||
   24363 GAGCCGTGCGTGAAATGGGAG...GCGGACGCCAATAAGCTGTACACCCT 24409

      70 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 87
         |   |||:::||||||||||||||||||||||   :::|||:::   |||
   24410 CTGCATGACCGATCCGGATGCGCCCAGTCGCAAGGATCCCAAGTTTAGGG 24459

      87 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 103
           |||:::||||||||||||   :::|||   |||:::|||:::   |||
   24460 AGTGGCACCATTGGCTGGTGGGCAACATACCCGGTGGAGATGTCGCCAAG 24509

     104 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 120
         |||::::::            ||||||||||||   :::   :::||||||
   24510 GGCGAGGTT............CTCTCCGCCTACGTGGGATCCGGGCCTCC 24547

     120 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 137
         |      :::|||:::|||||||||   |||:::::|||   |||:::
   24548 ACCAGACACCGGACTCCATCGTTACGTTTTCCTGATCTACGAGCAGCGGT 24597

     137 lyLysValIle.........SerLeuLeuProLysGluAsn...LysThr 149
           |||:::            :::   ||||||:::::::::
   24598 GCAAGCTCACATTCGACGAGAAGCGACTGCCCAATAACAGCGGAGATGGA 24647

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
         ||||||:::::|||:::   :::|||   :::::::::   |||||||::
   24648 CGCGGTGGCTTCAAAATCGCCGAGTTCGCCAAGAAGTACGCCCTCGGCAA 24697

     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAsp 180
         :|||   |||:::::::   ::::::::::::   |||:::|||
   24698 TCCCATTGCCGGTAACCTCTACCAGGCTGAG...TACGACGAC 24737

  . . . .
```

OM of: US-09-215-435-225_COPY_23_227 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 15, 2001 4:29 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2157/app_query.fasta_1.1519
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09215435_@CGN1_1_447 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 1043.840000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480	+	1115.00	2318.95	3.4e-121	681	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660	+	1115.00	2316.86	4.4e-121	826	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013	+	1115.00	2316.86	4.4e-121	826	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252	+	1115.00	2316.86	4.4e-121	826	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490	+	1115.00	2315.97	5.0e-121	897	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882	+	1111.00	2307.16	1.5e-120	932	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426	+	1109.00	2301.90	3.0e-120	1028		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539	+	1106.00	2296.94	5.7e-120	909	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582	+	1104.00	2292.81	9.7e-120	903	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431	+	1089.00	2260.76	5.9e-118	952	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350	+	1080.00	2242.61	6.1e-117	890	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582	+	543.00	1123.27	1.4e-54	447	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30001	+	237.50	472.99	2.2e-18	1047		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58685	+	237.50	472.99	2.2e-18	1047		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80734	+	205.00	405.28	1.3e-14	1001		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30002	+	205.00	401.29	2.2e-14	1447		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58686	+	205.00	401.29	2.2e-14	1447		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ14832	+	201.50	400.08	2.6e-14	822	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60684	+	182.00	363.96	2.7e-12	528	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC42966	+	182.00	363.96	2.7e-12	528	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60683	+	164.00	326.19	3.4e-10	528	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC48359	+	164.00	322.32	5.5e-10	754	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC34682	+	164.00	322.03	5.8e-10	775	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28490	+	164.00	320.95	6.6e-10	856	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ28491	-	164.00	320.95	6.6e-10	856	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT60139	+	148.00	286.49	5.5e-08	929	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90636	+	145.50	287.49	4.8e-08	522	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90668	+	145.00	286.51	5.5e-08	519	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT60140	+	140.50	274.33	2.6e-07	668	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC93828	+	121.00	239.85	2.2e-05	369	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA87678	+	90.50	163.06	0.4118	1200		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX39569	+	90.00	174.74	0.0921	371	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV63925	+	89.50	166.06	0.2803	750	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX81032	+	89.50	166.06	0.2803	750	!	
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85554	+	89.50	159.47	0.6521	1376		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA87700	+	88.00	158.19	0.7690	1159		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ01425	-	88.00	84.46	9.8e+03	10386		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF27717	+	87.50	143.64	4.97	4023		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ51531	+	87.50	141.74	6.34	4795		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX16969	+	87.50	141.74	6.34	4795		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC53434	+	86.00	153.24	1.45	124		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77512	+	84.50	137.00	11.65	415		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ43779	+	83.00	141.49	6.55	205		
/cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC51786	+	82.50	143.54	5.03	154		
/cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX13229	+	82.50	115.46	184.49	205		

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94480

seq_documentation_block:
ID AAF94480 standard; cDNA; 681 BP.
XX
AC AAF94480;
XX
DT 04-JUN-2001 (first entry)
XX
DE Human hydrophobic domain containing protein clone HP03880 cDNA #104.
XX
KW Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW behavioural characteristic; immune response; ss.
XX
OS Homo sapiens.
XX
PN WO200112660-A2.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-JP05356.
XX
PR 17-AUG-1999; 99JP-0230344.
PR 07-SEP-1999; 99JP-0252551.
PR 01-OCT-1999; 99JP-0281132.
PR 22-OCT-1999; 99JP-0301624.
PR 04-NOV-1999; 99JP-0313877.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2001-160059/16.
DR P-PSDB; AAB88590.
XX
PT Human proteins with hydrophobic domains and the DNAs which encode them
PT are useful for treating autoimmune disorders, burns and tumors and for
PT screening novel pharmaceuticals -
XX
PS Claim 3; Page 426-427; 518pp; English.
XX
CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC osteopathic, anti-inflammatory and cytostatic activities, and can be
CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC probes for genetic diagnosis and gene sources for gene therapy or for
CC producing (I) in large quantities. Cells containing (II) are used for
CC the detection of ligands or receptors corresponding to membrane or
CC secretory proteins and to screen small molecule novel pharmaceuticals.
CC Antibodies directed to (I) can be used for the detection, quantification
CC and purification of (I). Activities of (I) may include cytokine and cell
CC proliferation/differentiation function, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC wound healing, as nutritional sources or supplements e.g. as amino acid,
CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,

```
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 681 BP; 179 A; 186 C; 183 G; 133 T; 0 other;


alignment_scores:
           Quality: 1115.00              Length:    205
             Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAF94480   ..

 Align seg 1/1  to: AAF94480  from: 1  to: 681

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 116

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 166

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 216

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 266

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 316

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 366

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 416

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 466

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     467 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 516

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     517 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 566

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     567 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 616

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     617 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 666

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     667 GAGATAGCTGCCTGC 681

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX97660

seq_documentation_block:
```

```
ID   AAX97660 standard; DNA; 826 BP.
XX
AC   AAX97660;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 124.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; AAY35976.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 255; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
           Quality: 1115.00              Length:    205
             Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAX97660   ..

 Align seg 1/1  to: AAX97660  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180
```

```
 34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

 51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

 67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

 84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

201 GluIleAlaAlaCys 205
    |||||||||||||||
681 GAGATAGCTGCCTGC 695
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC00013

seq_documentation_block:

```
ID   AAC00013 standard; cDNA; 826 BP.
XX
AC   AAC00013;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein cDNA sequence #4.
XX
KW   Human; secreted protein; 5' EST; expressed sequence tag; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             15..698
FT                   /*tag=  a
FT                   /product= "secreted protein"
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00016.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Example 19; SEQ ID 7; 71pp + CD-ROM; English.
XX
CC   The present sequence is a full length cDNA encoding a human
CC   secreted protein. The cDNA was obtained from a 5' EST using first and
CC   second strand synthesis procedures. 5' ESTs were prepared from total
CC   human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region
CC   (UTR) of the mRNA because they are often obtained from oligo-dT primed
CC   cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC   sequences derived from the 5' ends of mRNAs and even in those cases where
CC   longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC   included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC   therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC   are also used in diagnostic, forensic, gene therapy and chromosome
CC   mapping procedures. They are used to obtain upstream regulatory sequences
CC   and to design expression and secretion vectors.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;
```

alignment_scores:

```
           Quality: 1115.00              Length:    205
             Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

```
 US-09-215-435-225_COPY_23_227 x AAC00013   ..

 Align seg 1/1  to: AAC00013  from: 1  to: 826

   1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

  17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

  34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

  51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

  67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

  84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

 101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

 117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480
```

```
    134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

    151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

    167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

    184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

    201 GluIleAlaAlaCys 205
        ||||||||||||||||
    681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42252

seq_documentation_block:
ID   AAZ42252 standard; cDNA; 826 BP.
XX
AC   AAZ42252;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human phosphatidylethanolamine-binding protein encoding cDNA.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   P-PSDB; AAY64647.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Example 21; Page 168-169; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 826 BP; 228 A; 229 C; 211 G; 158 T; 0 other;


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAZ42252   ..

 Align seg 1/1  to: AAZ42252  from: 1  to: 826

      1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

     17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

     34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

     51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

     67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

     84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

    101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

    117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

    134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

    151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

    167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

    184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

    201 GluIleAlaAlaCys 205
        ||||||||||||||||
    681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF94490
```

```
seq_documentation_block:
ID   AAF94490 standard; cDNA; 897 BP.
XX
AC   AAF94490;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Human hydrophobic domain containing protein clone HP03880 cDNA #114.
XX
KW   Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
KW   antianaemic; vulnerary; antiulcer; osteopathic; anti-inflammatory;
KW   cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
KW   HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
KW   inflammatory bowel disease; nutritional supplement; appetite; vaccine;
KW   behavioural characteristic; immune response; ss.
XX
OS   Homo sapiens.
XX
PN   WO200112660-A2.
XX
PD   22-FEB-2001.
XX
PF   10-AUG-2000; 2000WO-JP05356.
XX
PR   17-AUG-1999;   99JP-0230344.
PR   07-SEP-1999;   99JP-0252551.
PR   01-OCT-1999;   99JP-0281132.
PR   22-OCT-1999;   99JP-0301624.
PR   04-NOV-1999;   99JP-0313877.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-160059/16.
DR   P-PSDB; AAB88590.
XX
PT   Human proteins with hydrophobic domains and the DNAs which encode them
PT   are useful for treating autoimmune disorders, burns and tumors and for
PT   screening novel pharmaceuticals -
XX
PS   Claim 4; Page 442-444; 518pp; English.
XX
CC   AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC   AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC   anti-HIV, neuroprotective, antianaemic, vulnerary, antiulcer,
CC   osteopathic, anti-inflammatory and cytostatic activities, and can be
CC   used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC   to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC   probes for genetic diagnosis and gene sources for gene therapy or for
CC   producing (I) in large quantities. Cells containing (II) are used for
CC   the detection of ligands or receptors corresponding to membrane or
CC   secretory proteins and to screen small molecule novel pharmaceuticals.
CC   Antibodies directed to (I) can be used for the detection, quantification
CC   and purification of (I). Activities of (I) may include cytokine and cell
CC   proliferation/differentiation function, immune stimulating or suppressing
CC   activity, haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC   activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC   multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC   inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC   wound healing, as nutritional sources or supplements e.g. as amino acid,
CC   carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC   processing and utilisation of dietary fat, protein, carbohydrate,
CC   vitamins and minerals, to effect behavioural characteristics, to affect
CC   appetite, and can act as antigens in vaccines to raise an immune response
CC   to the protein or another material cross-reactive with the protein.
XX
SQ   Sequence 897 BP; 226 A; 253 C; 234 G; 184 T; 0 other;
```

```
alignment_scores:
           Quality: 1115.00              Length:    205
             Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAF94490   ..

 Align seg 1/1  to: AAF94490  from: 1  to: 897

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     165 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 214

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 264

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     265 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 314

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     315 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 364

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     365 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 414

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     415 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 464

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 514

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     515 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 564

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     565 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 614

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 664

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     665 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 714

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     715 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 764

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     765 GAGATAGCTGCCTGC 779

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC84882

seq_documentation_block:
ID   AAC84882 standard; cDNA; 932 BP.
XX
AC   AAC84882;
XX
DT   20-APR-2001  (first entry)
XX
DE   Human SEC1 nucleic acid sequence (clone ID 3445452).
```

```
XX
KW   SECX; cytostatic; gynecological; gene therapy; screening assay; human;
KW   chromosomal mapping; forensic biology; cell proliferation; cancer;
KW   cell differentiation; immune associated disorder; gestational disease;
KW   SEC1; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             113..781
FT                   /*tag=  a
FT                   /product= "SEC1"
FT   sig_peptide     113..181
FT                   /*tag=  b
FT   mat_peptide     182..778
FT                   /*tag=  c
XX
PN   WO200078802-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17328.
XX
PR   23-JUN-1999;   99US-0140584.
PR   20-JUL-1999;   99US-0144722.
PR   16-SEP-1999;   99US-0154520.
PR   22-JUN-2000; 2000US-0604286.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Vernet C,  Yang M,  Boldog FL;
PI   Herrmann JL;
XX
DR   WPI; 2001-071385/08.
DR   P-PSDB; AAB48368.
XX
PT   Polynucleotides encoding SECX proteins useful for treating disease
PT   characterized by an aberrant level of cell proliferation and/or
PT   differentiation like cancer or immune associated disorders -
XX
PS   Claim 3; Fig 1; 132pp; English.
XX
CC   The invention relates to human SECX polypeptides and polynucleotides
CC   encoding them. The SECX polypeptides can be expressed by standard
CC   recombinant methodology.  The SECX polypeptides are useful for treating
CC   or preventing a SECX-associated disorder. The invention is useful in
CC   screening assays; detection assays (e.g. chromosomal mapping, cell and
CC   tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC   prognostic assays, monitoring clinical trials, and pharmocogenomics); and
CC   methods of treatment (e.g. therapeutic and prophylactic), especially
CC   disorders characterized by aberrant cell proliferation and/ or
CC   differentiation like cancer or immune associated disorders or gestational
CC   disease. The present sequence represents a SEC1 nucleic acid sequence.
XX
SQ   Sequence 932 BP; 244 A; 257 C; 244 G; 187 T; 0 other;


alignment_scores:
            Quality: 1111.00              Length:    205
              Ratio:  5.420                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAC84882   ..

 Align seg 1/1  to: AAC84882  from: 1  to: 932

      1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    179 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 228

     17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    229 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 278

     34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    279 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 328

     51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    329 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 378

     67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    379 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 428

     84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    429 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 478

    101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
        ||||||:::||||||||||||||||||||||||||||||||||||||||||
    479 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 528

    117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    529 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 578

    134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    579 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 628

    151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    629 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 678

    167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    679 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 728

    184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    729 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 778

    201 GluIleAlaAlaCys 205
        ||||||||||||||||
    779 GAGATAGCTGCCTGC 793

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78426

seq_documentation_block:
ID   AAA78426 standard; cDNA; 1028 BP.
XX
AC   AAA78426;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:56.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
```

```
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24482.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 475; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1028 BP; 309 A; 268 C; 259 G; 186 T; 6 other;


alignment_scores:
           Quality: 1109.00              Length:    205
             Ratio:  5.436                Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAA78426   ..

 Align seg 1/1  to: AAA78426  from: 1  to: 1028

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     219 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 268

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     269 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 318

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     319 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 368

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     369 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 418

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     419 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 468

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     469 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 518

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     519 CTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 568

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     569 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 618

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     619 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 668

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     669 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 718

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     719 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 768

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     769 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 818

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     819 GAGATAGCTGCCTGC 833

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77539

seq_documentation_block:
ID   AAC77539 standard; cDNA; 909 BP.
XX
AC   AAC77539;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF3094 polynucleotide sequence SEQ ID NO:6187.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43330.
XX
```

```
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5369-5370; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 909 BP; 232 A; 254 C; 236 G; 186 T; 1 other;


alignment_scores:
            Quality: 1106.00              Length:    205
              Ratio:  5.422                Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAC77539   ..

 Align seg 1/1  to: AAC77539  from: 1  to: 909

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     166 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 215

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     216 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 265

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     266 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 315

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         |||||||||||||||||||||||||||||||||||   ||||||||
     316 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGTACGGCGCAAC 365

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     366 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 415

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     416 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 465

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     466 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 515

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     516 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 565

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     566 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 615

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     616 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 665

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     666 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 715

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     716 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 765

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     766 GAGATAGCTGCCTGC 780

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA15582

seq_documentation_block:
ID   AAA15582 standard; cDNA; 903 BP.
XX
AC   AAA15582;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human phospholipid binding protein 2, PLBP2 gene.
XX
KW   Human; phospholipid binding protein; PLBP2; foetal development disorder;
KW   reproduction disorder; cell proliferation disorder; immune response;
KW   autoimmune disorder; AIDS; infertility; cytostatic; immunosuppressive;
KW   gene therapy; hereditary neuropathy;
KW   phosphatidylethanolamine binding protein D1; PE-BP D1; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..771
FT                   /*tag= a
FT                   /product= "Human PLBP2"
XX
PN   US6063767-A.
XX
PD   16-MAY-2000.
XX
PF   09-DEC-1998;   98US-0208718.
XX
PR   28-OCT-1997;   97US-0958820.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Corley NC,  Shah P,  Lal P,  Hillman JL;
XX
DR   WPI; 2000-375529/32.
DR   P-PSDB; AAY94263.
XX
PT   New purified phospholipid binding proteins 1 and 2 useful for
PT   diagnosing, treating or preventing diseases disorders associated with
PT   fetal development, reproduction, cell proliferation, and the immune
PT   response  -
XX
PS   Example 5; Fig 2; 37pp; English.
XX
CC   The present sequence is the phospholipid binding protein 2 (PLBP2) gene.
CC   This gene is expressed in lung, prostate and heart tissues. Also, the
CC   protein is expressed in foetal tumour tissues. PLBP2 may be used for the
CC   diagnosis, prevention, or treatment of disorders associated with foetal
CC   development (e.g. hereditary neuropathies), reproduction (e.g.
CC   infertility), cell proliferation (e.g. cancers), and the immune response
CC   (AIDS). PLBP2 antibodies may also be developed for potential drug
```

```
CC   screening or to quantitate PLBP2 gene expression in biopsied tissues. The
CC   PLBP2 gene may be administered for gene therapy of disorders associated
CC   with PLBP2. PLBP2 has high homology with the phosphatidylethanolamine
CC   binding protein D1, PE-BP D1, of Onchocerca volvulus. PE-BP DI is thought
CC   to play a role in transport or signal mechanisms between membranes and
CC   the cytoplasm.
XX
SQ   Sequence 903 BP; 222 A; 251 C; 242 G; 188 T; 0 other;


alignment_scores:
            Quality: 1104.00              Length:    205
              Ratio:  5.412                Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.024

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAA15582   ..

 Align seg 1/1  to: AAA15582  from: 1  to: 903

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 203

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     204 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 253

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     254 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 303

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     304 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 353

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     354 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 403

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     404 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 453

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||:::||||||||||||||||||||||||||||||||||||||||||
     454 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 503

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     504 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 553

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     554 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 603

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     604 GGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACC 653

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
       . ||||||||||||||||||||||||||||||||||||||||||||||||||
     654 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 703

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         |||||||||||||||   ||||||||||||||||||||||||||||||||
     704 TCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 753

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     754 GAGATAGCTGCCTGC 768

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78431

seq_documentation_block:
ID   AAA78431 standard; cDNA; 952 BP.
XX
AC   AAA78431;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human secreted protein gene 46 SEQ ID NO:61.
XX
KW   Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW   antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW   antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
KW   neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW   immune system disorder; angiogenesis; hyperproliferative disorder;
KW   cardiovascular disorder; apoptosis; neurological disease;
KW   infectious disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200035937-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-US29950.
XX
PR   17-DEC-1998;   98US-0112809.
PR   18-DEC-1998;   98US-0113006.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ebner R,  Rosen CA,  Endress GA,  Soppet DR,  Ni J;
PI   Duan DR,  Moore PA,  Shi Y,  Lafleur DW,  Olsen HS,  Florence K;
XX
DR   WPI; 2000-431566/37.
DR   P-PSDB; AAB24487.
XX
PT   Forty seven human nucleic acids encoding secreted proteins, useful in
PT   the treatment, prevention and diagnosis of cancers, disorders of the
PT   immune system, angiogenesis disorders, neurological diseases and
PT   hyperproliferative disorders -
XX
PS   Claim 1; Page 478; 562pp; English.
XX
CC   The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC   human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: cytostatic; antianaemic;
CC   antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC   antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC   nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC   Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC   agonists may be useful in treating, preventing, and/or diagnosing other
CC   diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC   disorders of the immune system; (c) angiogenesis disorders; (d)
CC   hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC   associated with increase apoptosis; (g) neurological diseases; and
CC   (h) infectious diseases. They are also used to promote wound healing.
CC   AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 952 BP; 310 A; 247 C; 223 G; 172 T; 0 other;


alignment_scores:
            Quality: 1089.00              Length:    207
              Ratio:  5.312                Gaps:      2
 Percent Similarity:  99.034   Percent Identity: 99.034

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAA78431   ..

 Align seg 1/1  to: AAA78431  from: 1  to: 952
```

```
     1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   129 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 178

    17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   179 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 228

    34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   229 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 278

    51 SerTrpMetGlu.ProIleValLysPheProGlyAlaValAspGlyAlaT 67
       |||||||||||| |||||||||||||||||||||||||||||||||||||
   279 TCCTGGATGGAAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAA 328

    67 hrTyr.IleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
       ||||| |||||||||||||||||||||||||||||||||||||||||||||
   329 CCTATAATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACC 378

    83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaA 100
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   379 CAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCG 428

   100 spLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAla 116
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   429 ACCTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCT 478

   117 ProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTy 133
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   479 CCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTA 528

   133 rLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrA 150
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   529 TCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTC 578

   150 rgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGlu 166
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   579 GAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAA 628

   167 ProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProTh 183
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   629 CCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAAC 678

   183 rLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnA 200
       |||||||||||||||||||||||||||||||||||||||||||||||||||
   679 CCTCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGG 728

   200 laGluIleAlaAlaCys 205
       ||||||||||||||||
   729 CGGAGATAGCTGCCTGC 745
```

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA96350

seq_documentation_block:

```
ID   AAA96350 standard; cDNA; 890 BP.
XX
AC   AAA96350;
XX
DT   08-FEB-2001  (first entry)
XX
DE   cDNA encoding a novel polypeptide designated PRO4408.
XX
KW   Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW   PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW   PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW   PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW   insulinemia; kidney disorder; Bergers disease; nephropathy;
KW   Schonlein-Henoch purpura; celiac disease; dermatitis herpetiformis;
KW   Crohns disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             89..760
FT                   /*tag=  a
FT   sig_peptide     89..154
FT                   /*tag=  b
XX
PN   WO200056889-A2.
XX
PD   28-SEP-2000.
XX
PF   01-MAR-2000; 2000WO-US05601.
XX
PR   23-MAR-1999;   99US-0125774.
PR   23-MAR-1999;   99US-0125778.
PR   24-MAR-1999;   99US-0125826.
PR   31-MAR-1999;   99US-0127035.
PR   05-APR-1999;   99US-0127706.
PR   21-APR-1999;   99US-0130359.
PR   27-APR-1999;   99US-0131270.
PR   27-APR-1999;   99US-0131272.
PR   27-APR-1999;   99US-0131291.
PR   04-MAY-1999;   99US-0132371.
PR   04-MAY-1999;   99US-0132379.
PR   04-MAY-1999;   99US-0132383.
PR   25-MAY-1999;   99US-0135750.
PR   08-JUN-1999;   99US-0138166.
PR   20-JUL-1999;   99US-0144791.
PR   03-AUG-1999;   99US-0146970.
PR   09-DEC-1999;   99US-0170262.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Desnoyers L,  Eaton DL,  Goddard A,  Godowski PJ,  Gurney AL,  Pan J;
PI   Stewart TA,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2000-628263/60.
DR   P-PSDB; AAB18923.
XX
PT   Novel secreted and transmembrane polypeptides useful for diagnosing
PT   tumour in a mammal, for identifying agonists and antagonists of the
PT   polypeptide and for therapeutic use
XX
PS   Claim 2; Fig 29; 222pp; English.
XX
CC   The present sequence encodes a secreted or transmembrane polypeptide.
CC   The specification describes polypeptides designated PRO1484, PRO4334,
CC   PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC   PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC   PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC   useful for diagnosing tumour in a mammal. The polypeptides, their
CC   agonists and antagonists are useful treating a condition associated with
CC   expression or activity of the polypeptide. Conditions treated include
CC   obesity, diabetes or hyper-or hypo-insulinemia. The polypeptides are
CC   capable of inducing proliferation of mammalian kidney mesangial cells
CC   and are therefore useful for treating kidney disorders associated with
CC   decreased mesangial cell function such as Bergers disease or other
CC   nephropathies associated with Schonlein-Henoch purpura, celiac disease,
CC   dermatitis herpetiformis or Crohns disease. The nucleic acids may be used
CC   to generate transgenic animals for use in development and screening of
CC   therapeutically useful reagents and also for chromosome identification
CC   and tissue typing.
XX
SQ   Sequence 890 BP; 228 A; 246 C; 234 G; 182 T; 0 other;
```

```
alignment_scores:
            Quality: 1080.00              Length:    205
              Ratio:  5.294                Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAA96350    ..
```

```
Align seg 1/1  to: AAA96350  from: 1  to: 890

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 204

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 254

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 304

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 354

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     355 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 404

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     405 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 454

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     455 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 504

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     505 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 554

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     555 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 604

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     605 GGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACC 654

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     655 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 704

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         |||||||||||||   ||||||||||||||||||||   |||||||||
     705 TCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.AACCAGGCA 753

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     754 GAGATAGCTGCCTGC 768

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX40582

seq_documentation_block:
ID   AAX40582 standard; cDNA; 447 BP.
XX
AC   AAX40582;
XX
DT   18-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID No: 182.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906550-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01232.
XX
PR   01-AUG-1997;   97US-0905144.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153780/13.
DR   P-PSDB; AAY11860.
XX
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
XX
PS   Claim 1; Page 285; 675pp; English.
XX
CC   AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins expressed in prostate, and encode the proteins
CC   given in AAY11716 to AAY11993 respectively. The proteins given represent
CC   the signal peptide and an N-terminal fragment of a secreted protein. The
CC   nucleic acid sequences can be used for producing secreted human gene
CC   products. They can also be used to develop products for diagnosis and
CC   therapy. The proteins obtained may have cytokine activity, cell
CC   proliferation and differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptides can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 447 BP; 98 A; 110 C; 136 G; 99 T; 4 other;

alignment_scores:
           Quality:  543.00              Length:     99
             Ratio:  5.541                Gaps:      0
 Percent Similarity:  98.990   Percent Identity: 98.990

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAX40582   ..

 Align seg 1/1  to: AAX40582  from: 1  to: 447

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 200

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 250

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     251 TTGGCTGCAAGGTTGTTCCTGATTGTDACAACTACAGACAGAAGATCACC 300

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 350

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 400
```

```
     84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAla 99
        ||||||||||||||||||||||||||||||||||||||||||||||||
    401 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCC 447

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ30001

seq_documentation_block:
ID   AAQ30001 standard; cDNA to mRNA; 1047 BP.
XX
AC   AAQ30001;
XX
DT   18-MAR-1993  (first entry)
XX
DE   HCNP precursor gene #1.
XX
KW   Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;
KW   neonatal rat; expression vector; neurodegenerative disorder; dementia;
KW   Alzheimer's disease; Parkinson's disease; ss.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             26..589
FT                   /*tag=  a
FT   5'UTR           1..25
FT                   /*tag=  b
FT   3'UTR           587..1047
FT                   /*tag=  c
FT   polyA_signal    987..992
FT                   /*tag=  d
FT   misc_RNA        29..61
FT                   /*tag=  e
FT                   /label= Neurotrophic_activity_region
XX
PN   EP511816-A.
XX
PD   04-NOV-1992.
XX
PF   27-APR-1992;   92EP-0303800.
XX
PR   27-APR-1991;   91JP-0124688.
XX
PA   (OJIK/) OJIKA K.
PA   (SUMU ) SUMITOMO PHARM CO LTD.
PA   (YAMA/) YAMAMOTO M.
XX
PI   Agui H,  Fukushima N,  Irie T,  Kojima S,  Nishihara T;
PI   Ojika K,  Ono K,  Tohdoh N,  Tojo S,  Ueki Y;
XX
DR   WPI; 1992-367633/45.
DR   P-PSDB; AAR27897.
XX
PT   Neurotrophic peptide derivs. - used for treating neurological
PT   degenerative disorders, e.g. Alzheimer's disease or Parkinson's
PT   disease
XX
PS   Claim 3; Page 31-32; 57pp; English.
XX
CC   The seqence given is the hippocampal cholinergic neurotrophic peptide
CC   (HCNP) gene.  The active part of the protein encoded by this sequence
CC   is located in the first eleven amino acids at the N terminus.  This
CC   gene was obtained by preparing mRNA from the hippocampal tissue of
CC   neonatal rats, 12 days after birth.  The mRNA is converted to cDNA by
CC   standard methods.  The cDNA can be used in expression vectors to
CC   produce both the full length protein and the active peptide fragment.
CC   The gene and the peptide are useful for the treatment of
CC   neurodegenerative disorders and dementia eg. Alzheimer's and
CC   Parkinson's disease.
XX
SQ   Sequence 1047 BP; 257 A; 296 C; 290 G; 204 T; 0 other;

alignment_scores:
           Quality:  237.50              Length:    185
             Ratio:  2.284                 Gaps:      6
 Percent Similarity:  56.216   Percent Identity:  32.973

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAQ30001   ..

 Align seg 1/1  to: AAQ30001  from: 1  to: 1047

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         |||||||||:::::         :::::::|||                ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      46 gGlnLysIleThrSerTrpMetGluProIleValLysProProGlyAlaV 63
         |   :::      ||||||                              |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
         .    |||      |||   ||||||:::::|||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::|||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||:::::::::|||  :::             ||||||
     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 338

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
           |||   :::   :::|||||||       :::|||:::|||||||||
     339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

     130 PhePheValTyrLeuGlnGluGlyLysValIleSer.............Le 142
         ::::::||||||   ||||||      :::                 ||
     389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

     142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
         |      |||::::::   :::|||||:::::|||:::::   |||
     439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

     159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

     176 GlnAsnTyrGlnAspSer.............................. 181
         :::   :::::||||||
     539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585

     182 ..............................ProThrLeuGlnAlaP 187
                                      ||||||:::|||:::
     586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

     187 roArg 188
           |||
     636 GTCGT 640

 seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ58685

 seq_documentation_block:
 ID   AAQ58685 standard; cDNA to mRNA; 1047 BP.
 XX
 AC   AAQ58685;
 XX
 DT   13-OCT-1994  (first entry)
 XX
 DE   Rat hippocampal cholinergic neurotrophic peptide precursor cDNA.
 XX
 KW   Rat hippocampal cholinergic neurotrophic peptide; rat HCNP;
 KW   nerve degeneration; acetylcholine synthesis; neurostimulation; ds.
 XX
 OS   Rattus norvegicus (Wistar).
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             26..589
FT                   /*tag= a
FT                   /product= rat_HCNP
FT   5'UTR           1..25
FT                   /*tag= b
FT   3'UTR           590..1047
FT                   /*tag= c
FT   polyA_signal    987..992
FT                   /*tag= d
FT   polyA_site      1008..1047
FT                   /*tag= e
XX
PN   WO9405788-A.
XX
PD   17-MAR-1994.
XX
PF   27-AUG-1993;   93WO-JP01214.
XX
PR   29-AUG-1992;   92JP-0253734.
PR   28-AUG-1992;   92JP-0254170.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
PA   (YAMA/) YAMAMOTO M.
XX
PI   Agui H,  Fukushima N,  Irie T,  Kamikawa Y,  Kojima S;
PI   Nishihara T,  Ojika K,  Ono K,  Taiji M,  Tohdoh N;
PI   Tojo S,  Ueki Y;
XX
DR   WPI; 1994-101193/12.
DR   P-PSDB; AAR49942.
XX
PT   Neurotrophic peptide(s), precursors and genes - used to treat
PT   nervous degeneration, increases acetylcholine synthesis
XX
PS   Claim 3; Page 153-154; 200pp; Japanese.
XX
CC   The first 11 N-terminal amino acids of the rat hippocampal
CC   cholinergic neurotrophic peptide precursor (encoded by AAQ58685)
CC   have neurotrophic activity and increase acetylcholine synthesis.
XX
SQ   Sequence 1047 BP; 257 A; 297 C; 289 G; 204 T; 0 other;


alignment_scores:
            Quality:  237.50            Length:    185
              Ratio:  2.284               Gaps:      6
 Percent Similarity:  56.216   Percent Identity: 32.973

alignment_block:
 US-09-215-435-225_COPY_23_227 x AAQ58685   ..

 Align seg 1/1  to: AAQ58685  from: 1  to: 1047

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         ||||||||||:::::         ::::::|||               ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
         |   :::     ||||||                    |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
            |||      |||   ||||||::::::||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||::::::::::|||   :::            ||||||

     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC...........CTCTCCG 338

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
           |||   :::   :::||||||      :::|||:::|||||||||
     339 AATACGTGGGCTCCGGACCTCCCAAAGACACAGGTCTGCACCGCTACGTC 388

     130 PhePheValTyrLeuGlnGluGlyLysValIleSer...........Le 142
         ::::::||||||   ||||||      :::                  ||
     389 TGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCATCCT 438

     142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
         |      |||::::::   :::||||||:::::::|||:::::::   |||
     439 CAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCTTCC 488

     159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
           :::::::::|||||||||   |||   |||:::|||   |||::::::
     489 GCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAGGCA 538

     176 GlnAsnTyrGlnAspSer.............................. 181
         :::   ::::::||||||
     539 GAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGCTGGCTGGGAA 585

     182 ..................................ProThrLeuGlnAlaP 187
                                          ||||||:::|||:::
     586 GTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGTCAA 635

     187 roArg 188
           |||
     636 GTCGT 640

seq_name: /cgn1_9/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80734

seq_documentation_block:
ID   AAQ80734 standard; cDNA to mRNA; 1001 BP.
XX
AC   AAQ80734;
XX
DT   04-JUL-1995  (first entry)
XX
DE   Phosphatidylethanolamine binding protein coding sequence.
XX
KW   Human; phosphatidylethanolamine binding protein; hPEBP; cDNA library;
KW   glioblastoma; T98G; aplasia; proliferation; glia; neuron;
KW   hematopoietic cell; immunological activity; neurological activity;
KW   inflammatory disease; tumour; antibody; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             53..616
FT                   /*tag= a
FT                   /product= hPEBP
XX
PN   EP628631-A.
XX
PD   14-DEC-1994.
XX
PF   30-MAY-1994;   94EP-0108321.
XX
PR   08-JUN-1993;   93JP-0137042.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Naito T,  Nakade S;
XX
DR   WPI; 1995-015698/03.
DR   P-PSDB; AAR64268.
XX
PT   New human phosphatidylethanolamine binding protein gene - used to
PT   develop prods. for diagnosis and treatment of diseases involving
PT   glias, neurons and other cells.
XX
PS   Claim 4; Page 13-14; 17pp; English.
```

XX
CC This sequence encodes the human phosphatidylethanolamine binding
CC protein, hPEBP. This cDNA was isolated from a cDNA library prepared
CC using RNA extracted from the human glioblastoma cell line T98G. The
CC protein can be used for the prevention of or in the treatment of
CC aplasia or abnormal proliferation of glia, neurons or hematopoietic
CC cells, depression or enhancement of immunological or neurological
CC activity, inflammatory disease, tumours or diseases induced by
CC abnormal lipid metabolism. Anti-PEBP antibodies may be used for
CC diagnosis of genetic diseases.
XX
SQ Sequence 1001 BP; 250 A; 249 C; 279 G; 223 T; 0 other;

alignment_scores:
Quality: 205.00 Length: 151
Ratio: 2.228 Gaps: 4
Percent Similarity: 60.927 Percent Identity: 33.113

alignment_block:
US-09-215-435-225_COPY_23_227 x AAQ80734 ..

Align seg 1/1 to: AAQ80734 from: 1 to: 1001

```
     30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
        |||||||||:::::          :::::::|||        :::   ||
    158 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTTAAGAATAG 198

     46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
        |            ||||||                          |||
    199 ACCCACCAGCATTTCGTGG.....................GATGGTCTTG 227

     63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
          :::|||      |||   ||||||:::::|||||||||||||||||||
    228 ATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCCGGATGCTCCCAGC 277

     80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
        |||   :::|||:::   |||   |||:::|||:::||||||:::::::
    278 AGGAAGGATCCCAAATACAGAGAATGGCATCATTTCCTGGTGGTCAACAT 327

     96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
        :||||||   |||:::::::::||   :::             ||||||
    328 GAAGGGCAATGACATCAGCAGTGGCACAGTC...........CTCTCCG 365

    113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
          |||   :::   :::|||||||      :::|||:::|||||||||
    366 ATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCTCCACCGCTATGTC 415

    130 PhePheValTyrLeuGlnGluGlyLysValIleSer...........Le 142
        ::::::|||||   |||:::      :::                   ||
    416 TGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTGACGAGCCCATCCT 465

    142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
        |      :::::::::      ||||||::::::|||:::      |||
    466 CAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAGGTGGCGTCCTTCC 515

    159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
          :::::::::::|||      |||   |||:::|||   :::::::::
    516 GTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCACGTGTTACCAGGCC 565

    176 Gln 176
        :::
    566 GAG 568
```

```
OM of: US-09-215-435-225_COPY_23_227 to: Issued_Patents_NA:*   out_format : pfs

Date: Nov 15, 2001  5:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2141/app_query.fasta_1.1519
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_161 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 228.030000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4 + 1104.00 2458.21  2.8e-129  903  !  S
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-3 +  212.50  448.08  2.6e-17   1047  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-14 +  205.00  427.44  3.7e-16   1447  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-644-372-1 +  201.50  426.10  4.4e-16   822  !  S
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-1 +  164.00  341.17  2.4e-11   855  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-3 -  164.00  341.15  2.4e-11   856  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-1021 -  122.50  249.57  3.0e-06   726  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-991 +  115.00  233.23  2.4e-05   692  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2 -  105.00  109.29  194.62    4403765
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-1 -  105.00  109.26  195.13    4411529
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-47 +   87.50  171.14  0.0698    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-11 +   87.50  171.14  0.0698    700  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-47 +   87.50  171.14  0.0698    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-8 +   87.50  148.86  1.22      4794  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-45 +   87.50  148.86  1.22      4795  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-45 +   87.50  148.86  1.22      4795  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-46 +   82.50  159.88  0.2960    700  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-46 +   82.50  159.88  0.2960    700  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-10 +   82.00  158.77  0.3413    699  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-07-765-830A-5 +   80.00  133.34  8.90      4258  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-056-105-10 +   77.50  126.09  22.54     4895  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-056-105-9 +   77.50  116.21  80.10     11495  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-288-630-1 +   77.00  140.03  3.77      1333  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-08-814-095-7 -   77.00  102.17  484.84    35060  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-697-610-1 +   76.00  131.76  10.89     2240  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-349-357-1 +   76.00  131.76  10.89     2240  !
/cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-16980-1 +   76.00  131.16  11.76     2359
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-211-417-2 +   76.00  127.33  19.24     3286  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-770-379-19 +   76.00   99.90  648.37    35100  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-757-669A-19 +   76.00   99.90  648.37    35100
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-791-115B-8 -   75.00  136.20  6.17      1257  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-791-115B-2 -   75.00  125.53  24.24     3160  !
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-368-590-1 -   74.00  112.79  124.09    7812  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-812-13 +   74.00  110.51  166.31    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-920-827-13 +   74.00  110.51  166.31    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-921-177-13 +   74.00  110.51  166.31    9515  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-362-577C-13 +   74.00  110.51  166.31    9515  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-920-828-13 +   74.00  110.51  166.31    9515  !
/cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-492-027A-9 +   73.50  127.83  18.05     1935  !
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-785-241-2 +   73.50  122.63  35.14     3031  !
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-189-760-5 +   73.00  129.43  14.70     1529
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-188-811-5 +   73.00  129.43  14.70     1529
/cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-514-422-5 +   73.00  129.43  14.70     1529
/cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-656-984A-33 +   73.00  128.29  17.01     1687
/cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-188-811-1 +   73.00  128.17  17.26     1704

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-208-718-4

seq_documentation_block:
; Sequence 4, Application US/09208718
; Patent No. 6063767
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer
;    APPLICANT:  Corley, Neil
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN PHOSPHOLIPID BINDING PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,718
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/958,820
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0379 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  LUNGTUT12
;      CLONE:  3126479
US-09-208-718-4


alignment_scores:
            Quality: 1104.00              Length:    205
              Ratio:  5.412                Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.024

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-208-718-4   ..

 Align seg 1/1  to: US-09-208-718-4  from: 1  to: 903

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 203

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     204 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 253
```

```
     34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    254 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 303

     51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    304 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 353

     67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    354 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 403

     84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    404 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 453

    101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
        ||||||:::||||||||||||||||||||||||||||||||||||||||||
    454 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 503

    117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    504 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 553

    134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    554 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 603

    151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    604 GGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACC 653

    167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    654 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 703

    184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
        |||||||||||||||   ||||||||||||||||||||||||||||||||
    704 TCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 753

    201 GluIleAlaAlaCys 205
        ||||||||||||||||
    754 GAGATAGCTGCCTGC 768

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-3

seq_documentation_block:
; Sequence 3, Application US/08403378B
; Patent No. 5759991
;  GENERAL INFORMATION:
;    APPLICANT:  TOHDOH, NAOKI
;    APPLICANT:  TOJO, SHIN-ICHIRO
;    APPLICANT:  KOJIMA, SHIN-ICHI
;    APPLICANT:  UEKI, YASUYUKI
;    APPLICANT:  NISHIHARA, TOSHIO
;    APPLICANT:  FUKUSHIMA, NOBUYUKI
;    APPLICANT:  IRIE, TSUNEMASA
;    APPLICANT:  ONO, KEIICHI
;    APPLICANT:  AGUI, HIDEO
;    APPLICANT:  OJIKA, KOSEI
;    TITLE OF INVENTION:  NEUROTROPHIC PEPTIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 PENNSYLVANIA AVENUE, NW
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,378B
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-124688
;      FILING DATE:  27-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-080398
;      FILING DATE:  30-MAR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-280590
;      FILING DATE:  27-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-333241
;      FILING DATE:  21-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-243003
;      FILING DATE:  12-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/758,043
;      FILING DATE:  12-SEP-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/873,764
;      FILING DATE:  27-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01214
;      FILING DATE:  27-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGART, WADDELL A
;      REGISTRATION NUMBER:  24,861
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1047 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus norvegicus
;      STRAIN:  Wistar
;      TISSUE TYPE:  hippocampal tissue of brain
;    IMMEDIATE SOURCE:
;      CLONE:  A010-12
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  26..586
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..25
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  26..586
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  587..1047
;    FEATURE:
;      NAME/KEY:  polyA_site
;      LOCATION:  1008..1047
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  987..992
US-08-403-378B-3
```

```
alignment_scores:
           Quality:  212.50              Length:    186
             Ratio:  2.063                Gaps:      8
 Percent Similarity:  55.376   Percent Identity: 32.796

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-403-378B-3   ..

 Align seg 1/1  to: US-08-403-378B-3  from: 1  to: 1047

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         |||||||||::::::        ::::::|||              ||
     131 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTCATGAATAG 171

      46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
         |   :::      ||||||                        |||
     172 ACCAAGCAGCATTTCATGG.....................GATGGCCTTG 200

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
            |||      |||   ||||||::::::||||||||||||||||||
     201 ATCCTGGGAAGCTCTACACCCTGGTCCTCACAGACCCCGATGCTCCCAGC 250

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     251 AGGAAGGACCCCAAATTCAGGGAGTGGCACCACTTCCTGGTGGTCAACAT 300

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||:::::::::|||   :::             ||||||
     301 GAAGGGCAACGACATTAGCAGTGGCACTGTC............CTCTCCG 338

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHis...ArgTyr 128
           |||   :::   |||   |||      :::|||:::|||   ||||||
     339 AATAC...GTGGGCTCCGGACCTAAAGACACAGGTCTGCACCGCCGCTAC 385

     129 GlnPheValTyrLeuGlnGluGlyLysValIleSer..... 141
         ::::::||||||   ||||||      :::
     386 GTCTGGCTGGTGTATGAGCAGGAGCAGCCTCTGAACTGTGACGAGCCCAT 435

     142 .LeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgP 158
          |||      |||::::::   :::||||||:::::::|||::::::   |
     436 CCTCAGCAACAAGTCTGGAGACAACCGCGGCAAGTTCAAGGTGGAGTCCT 485

     158 heLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMet 174
         ||   :::::::::|||||||||   |||   |||:::|||   |||:::
     486 TCCGCAAGAAGTACCACCTGGGAGCCCCGGTGGCCGGCACGTGCTTCCAG 535

     175 ThrGlnAsnTyrGlnAspSer..... 181
         ::::::   ::::::||||||
     536 GCAGAG...TGGGATGACTCTGTGCCCAAGCTGCATGATCAGTCGGCTGG 582

     182 .....ProThrLeuGlnA 186
                           ||||||:::|||:
     583 GAAGTAGGGGCGCTGCAGAGCCCGCAGCCCCGGGGACCCCACAGTACAGT 632

     186 laProArg 188
         ::   |||
     633 CAAGTCGT 640

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-08-403-378B-14

seq_documentation_block:
; Sequence 14, Application US/08403378B
; Patent No. 5759991
;  GENERAL INFORMATION:
;    APPLICANT:  TOHDOH, NAOKI
;    APPLICANT:  TOJO, SHIN-ICHIRO
;    APPLICANT:  KOJIMA, SHIN-ICHI
;    APPLICANT:  UEKI, YASUYUKI
;    APPLICANT:  NISHIHARA, TOSHIO
;    APPLICANT:  FUKUSHIMA, NOBUYUKI
;    APPLICANT:  IRIE, TSUNEMASA
;    APPLICANT:  ONO, KEIICHI
;    APPLICANT:  AGUI, HIDEO
;    APPLICANT:  OJIKA, KOSEI
;    TITLE OF INVENTION:  NEUROTROPHIC PEPTIDE DERIVATIVES
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 PENNSYLVANIA AVENUE, NW
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,378B
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-124688
;      FILING DATE:  27-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-080398
;      FILING DATE:  30-MAR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-280590
;      FILING DATE:  27-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-333241
;      FILING DATE:  21-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-243003
;      FILING DATE:  12-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/758,043
;      FILING DATE:  12-SEP-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/873,764
;      FILING DATE:  27-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/01214
;      FILING DATE:  27-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIGGART, WADDELL A
;      REGISTRATION NUMBER:  24,861
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  placental tissue
;    IMMEDIATE SOURCE:
;      CLONE:  p1-3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  120..680
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..119
;    FEATURE:
```

```
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..680
;     FEATURE:
;       NAME/KEY:  3'UTR
;       LOCATION:  681..1447
US-08-403-378B-14


alignment_scores:
            Quality:  205.00              Length:    151
              Ratio:  2.228                 Gaps:      4
 Percent Similarity:  60.927   Percent Identity: 33.113

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-403-378B-14    ..

 Align seg 1/1  to: US-08-403-378B-14  from: 1  to: 1447

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         |||||||||:::::          ::::::|||          :::   ||
     225 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTTAAGAATAG 265

      46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
         |           ||||||                           |||
     266 ACCCACCAGCATTTCGTGG.....................GATGGTCTTG 294

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
         :::|||      |||   ||||||:::::|||||||||||||||||||
     295 ATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCCGGATGCTCCCAGC 344

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::|||||||:::::::
     345 AGGAAGGATCCCAAATACAGAGAATGGCATCATTTCCTGGTGGTCAACAT 394

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||:::::::::|||   :::            ||||||
     395 GAAGGGCAATGACATCAGCAGTGGCACAGTC...........CTCTCCG 432

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
         |||   :::   :::|||||||      :::|||:::|||||||||
     433 ATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCTCCACCGCTATGTC 482

     130 PhePheValTyrLeuGlnGluGlyLysValIleSer...........Le 142
         ::::::|||||   |||:::    :::                   ||
     483 TGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTGACGAGCCCATCCT 532

     142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
         |      :::::::::      ||||||:::::|||:::      |||
     533 CAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAGGTGGCGTCCTTCC 582

     159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
         :::::::::::|||      |||   |||:::|||   :::::::::
     583 GTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCACGTGTTACCAGGCC 632

     176 Gln 176
         :::
     633 GAG 635

 seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-644-372-1

 seq_documentation_block:
 ; Sequence 1, Application US/07644372
 ; Patent No. 5416009
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Lazzeri, Mario E.
 ;    APPLICANT:  Nutman, Thomas B.
 ;    APPLICANT:  Weiss, Niklaus
 ;    TITLE OF INVENTION:  A DNA SEGMENT ENCODING A SPECIFIC
 ;    TITLE OF INVENTION:  IMMUNODIAGNOSTIC ANTIGEN
 ;    NUMBER OF SEQUENCES:  2
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
 ;      STREET:  1615 L. Street, N.W.
 ;      CITY:  Washington
 ;      STATE:  D.C.
 ;      COUNTRY:  USA
 ;      ZIP:  20036
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/07/644,372
 ;      FILING DATE:  19910123
 ;      CLASSIFICATION:  435
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  (202)861-3000
 ;      TELEFAX:  (202)822-0944
 ;  INFORMATION FOR SEQ ID NO:  1:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  822 base pairs
 ;      TYPE:  NUCLEIC ACID
 ;      STRANDEDNESS:  double
 ;      TOPOLOGY:  linear
 ;    FEATURE:
 ;      NAME/KEY:  CDS
 ;      LOCATION:  52..507
 US-07-644-372-1


 alignment_scores:
             Quality:  201.50              Length:    147
               Ratio:  2.144                 Gaps:      7
  Percent Similarity:  63.946   Percent Identity: 36.054

 alignment_block:
  US-09-215-435-225_COPY_23_227 x US-07-644-372-1    ..

  Align seg 1/1  to: US-07-644-372-1  from: 1  to: 822

       33 AsnIleGlyCysLysValValPro...AspCysAsnAsnTyrArgGlnLy 48
          |||:::|||   ||||||||||||   :::   :::|||          ||
      244 AATCTGGGCAATGAACTTACGCCGACGCAGGTAAAGAATCAGCCGACAAA 293

       48 sIleThrSerTrpMetGluProIleValLysPheProGlyAlaValAspG 65
          |:::   ||||||                                |||     |
      294 AGTA...TCATGG.....................GATGCGGAACCTG 316

       65 lyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAla 81
          |||||   |||   |||||||||:::|||||||||||||||||||||||
      317 GAGCCTTATATACGCTCGTTATGACTGATCCGGACGCACCATCTCGAAAA 366

       82 GluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGl 98
          :::|||      |||   |||:::|||||||||:::   :::|||:::||
      367 AACCCCGTATTCAGAGAGTGGCACCATTGGTTGATAATTAATATTTCTGG 416

       98 yAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr. 114
          |   ::::::::::::|||   :::            ||||||   |||
      417 ACAAAATGTTAGCAGTGGCACAGTG...........TTATCTGATTATT 454

      115 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 131
          :::   :::|||          :::|||:::|||||||||   |||::
      455 GGATCA...GGTCCACGAAAAGGCACAGGACTTCATCGTTATGTATTCTT 501

      131 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 148
          :||||||   |||   |||:::::::   :::          :::::::::
      502 GGTTTATAAACAACCTGGAAGTATCACGGAT......ACTCAACATGGCG 545

      148 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 164
          :::|||   :::::::|||:::      |||   |||:::   ||||||
      546 GAAATCGCCGAAATTTCAAAGTTATGGATTTTGCAAACAAACATCACTTG 595

      165 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsn 177
          |||:::|||   |||:::::::   |||:::::::::::
      596 GGAAATCCAGTTGCCGGAAACTTCTTCCAGGCTAAACAT 634
```

```
seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-1

seq_documentation_block:
; Sequence 1, Application US/09060726A
; Patent No. 6225530
; GENERAL INFORMATION:
;  APPLICANT:  Weigel, Detlef
;  APPLICANT:  Salk Institute
;  TITLE OF INVENTION:  FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE:  SALKINS.026A
;  CURRENT APPLICATION NUMBER: US/09/060,726A
;  CURRENT FILING DATE:  1998-04-15
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 855
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (63)...(590)
US-09-060-726A-1


alignment_scores:
           Quality:  164.00            Length:    113
             Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-060-726A-1   ..

 Align seg 1/1  to: US-09-060-726A-1  from: 1  to: 855

      68 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 84
         |||   |||||||||||||||||||||:::||||||   ::::::|||:: 
     255 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 304

      84 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 101
         :   |||   :::   ||||||||||||||||||||   ::::::      
     305 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 350

     101 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 117
               |||         |||:::|||:::   :::|||:::   |||  
     351 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 395

     118 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 134
         ||||||   |||      |||:::||||||      |||:::::::::   
     396 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 439

     134 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 151
          |||   |||:::               |||                   
     440 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 470

     151 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 163
           :::|||:::::::::           :::|||   :::   ::::::  
     471 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 518

     164 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 176
         ||||||   |||   |||::::::   :::      |||  
     519 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 557

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-060-726A-3

seq_documentation_block:
; Sequence 3, Application US/09060726A
; Patent No. 6225530
; GENERAL INFORMATION:
;  APPLICANT:  Weigel, Detlef
;  APPLICANT:  Salk Institute
;  TITLE OF INVENTION:  FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE:  SALKINS.026A
;  CURRENT APPLICATION NUMBER: US/09/060,726A
;  CURRENT FILING DATE:  1998-04-15
;  NUMBER OF SEQ ID NOS:  13
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-060-726A-3


alignment_scores:
           Quality:  164.00            Length:    113
             Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-060-726A-3/rev   ..

 Align seg 1/1  to reverse of: US-09-060-726A-3  from: 1  to: 856

      68 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 84
         |||   |||||||||||||||||||||:::||||||   ::::::|||:: 
     602 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 553

      84 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 101
         :   |||   :::   ||||||||||||||||||||   ::::::      
     552 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 507

     101 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 117
               |||         |||:::|||:::   :::|||:::   |||  
     506 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 462

     118 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 134
         ||||||   |||      |||:::||||||      |||:::::::::   
     461 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 418

     134 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 151
          |||   |||:::               |||                   
     417 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 387

     151 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 163
           :::|||:::::::::           :::|||   :::   ::::::  
     386 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 339

     164 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 176
         ||||||   |||   |||::::::   :::      |||  
     338 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 300

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-1021

seq_documentation_block:
; Sequence 1021, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
```

```
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1021:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  726 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1025UP
US-08-998-416-1021


alignment_scores:
            Quality:  122.50              Length:     89
              Ratio:  2.149                Gaps:      3
 Percent Similarity:  64.045   Percent Identity: 34.831

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-998-416-1021/rev   ..

 Align seg 1/1  to reverse of: US-08-998-416-1021  from: 1  to: 726

      65 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 81
         |||       :::   |||   |||:::||||||||||||||||||||::
     717 GGGGACCTATTCACGCTGGGGATGACAGACCCAGATGCTCCCTCGCGGTC 668

      81 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys. 97
         ::::   :::       :::   |||:::|||   |||:::|||
     667 GGACCACAAGTGGTCGGAATACTGCCACTTTCTGGAAACGAACATAACGC 618

      98 ..GlyAlaAsp......LeuLysLysGlyLysIleGlnGlyGlnGluLeu 111
           |||:::|||      ::::::         :::::::|||
     617 TGGGCTCGGATGACGGGGTGTCGCACGTGGTGCTAAAGGGCACCCCGCAG 568

     112 SerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTy 128
           ::::::::::|||:::|||||||||   :::|||   ||||||||
     567 GTGGAGCACATGGGCCCTGCGCCGCCGGCCGGCACAGGGGCTCACCGGTA 518

     128 rGlnPheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProL 145
         |   ::::::::::::   |||   |||::::::         |||   :
     517 CGTGTGGTTGTTGTTTCGGCAGCCCGGGCGGCTA......GAACTGAGCG 474

     145 ysGluAsnLysThrArg 150
         ::|||:::   ||||||
     473 AGGAGGAGGTCACCCGC 457

 seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-08-998-416-991

 seq_documentation_block:
 ; Sequence 991, Application US/08998416
 ; Patent No. 6239264
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Philippsen, Peter
 ;    APPLICANT:  Pohlmann, Rainer
 ;    APPLICANT:  Steiner, Sabine
 ;    APPLICANT:  Mohr, Christine
 ;    APPLICANT:  Wendland, Jurgen
 ;    APPLICANT:  Knechtle, Philipp
 ;    APPLICANT:  Rebischung, Corinne
 ;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
 ;    TITLE OF INVENTION:  AND USES THEREOF
 ;    NUMBER OF SEQUENCES:  1152
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  No. 6239264artis Corporation
 ;      STREET:  3054 Cornwallis Road
 ;      CITY:  Research Triangle Park
 ;      STATE:  No. 6239264th Carolina
 ;      COUNTRY:  USA
 ;      ZIP:  27709
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/998,416
 ;      FILING DATE:  24-DEC-1997
 ;      CLASSIFICATION:  435
 ;    PRIOR APPLICATION DATA:
 ;      APPLICATION NUMBER:  CH 0016/97
 ;      FILING DATE:  31-DEC-1996
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Meigs, J. Timothy
 ;      REGISTRATION NUMBER:  38,241
 ;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  919-541-8587
 ;      TELEFAX:  919-541-8689
 ;  INFORMATION FOR SEQ ID NO:  991:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  692 base pairs
 ;      TYPE:  nucleic acid
 ;      STRANDEDNESS:  single
 ;      TOPOLOGY:  linear
 ;    MOLECULE TYPE:  DNA (genomic)
 ;    ORIGINAL SOURCE:
 ;      ORGANISM:  PAG1609RP
 US-08-998-416-991


 alignment_scores:
             Quality:  115.00              Length:    140
               Ratio:  1.620                Gaps:      8
  Percent Similarity:  50.714   Percent Identity: 30.000

 alignment_block:
  US-09-215-435-225_COPY_23_227 x US-08-998-416-991   ..

  Align seg 1/1  to: US-08-998-416-991  from: 1  to: 692

       4 GluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAspThrLeuPh 20
         |||||||::|||   :::|||
     336 GAAAACGGTCCATCAGGACAC.............................. 356

      20 eCysGlnGlyLeuGluValPheTyrPro.................GluL 31
                 |||   |||   ||||||                      |
     357 .........CTTGTAGTGGAGTACCCGGGCGAGTCTACAGCGGTGACGC 396

      31 euGlyAsnIle............GlyCysLysValValProAspCysAsn 43
         |||||||||:::           :::   ::::::||||||:::
     397 TGGGGAACGTTATGCCTGTGGAGGCTACGCAGACGGTGCCCAAC...... 440

      44 AsnTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPhePr 60
                                     :::      |||:::      ||
     441 .........................CTGATGTTAATCACGACCGAGCC 463
```

```
      60 oGlyAlaVal...AspGlyAlaThrTyrIleLeuValMetValAspProA 76
         ||||   |||   :::|||      :::   |||:::|||:::||||||| 
     464 GGGAATCGTCAGGGAGGGGGACCTATTCACGCTGGCGATGACAGACCCAG 513

      76 spAlaProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeu 92
         ||||||||||||||:::::   :::         :::   |||:::||| 
     514 ATGCTCCCTCGCGGTCGGACCACAAGTGGTCGGAATACTGCCACTTTCTG 563

      93 ValThrAspIleLys...GlyAlaAsp......LeuLysLysGlyLysIl 106
            |||:::|||      |||:::|||      ::::::         :: 
     564 GAAACGAACATAACGCTGGGCTCGGATGACGGGGTGTCGCACGTGGTGCT 613

     106 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPro.AlaHis 122
         ::::|||               ::::::::::|||:::|||||| ||| 
     614 AAAGGGCACCCCGCAGGTGGAGCACATGGGCCCTGCGCCGCCGGGCCGGC 663

     123 SerGlyPheHisArgTyr 128
         :::|||   ||||||||| 
     664 ACAGGGGCTCACCGGTAC 681

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
           Quality:  105.00            Length:     93
             Ratio:  2.283               Gaps:      6
 Percent Similarity:  49.462   Percent Identity: 34.409

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-103-840A-2/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-2  from: 1  to: 4403765

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         :::|||      |||         |||   ||||||
  2154305ACCTGGTCGGCGCCG.........TTTGGCGGCGCA.............. 2154279

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
            ||||||:::   ||||||||||||||         ||||||
  2154278.......CTCGTTGTCGATGATCCGGACGCACCT......CGCGAACCT. 2154243

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys......Gly 98
                 :::   ||||||:::|||      |||         ::: 
  2154242...........TACGTCCATTGGATCGTGATCGGGATCGCCCCTGGTGCT 2154204

      99 AlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSer........ 112
         ::::::        |||:::      |||      :::|||        .
  2154203GGCAGCACCGCCGATGGTGAGACTCCCGGTGGCGGAATCAGCCTGCCGAA 2154154

     113 ..............AlaTyrGlnAlaProSerProProAlaHisSerG 124
                       ||||||   :::|||   |||||||||   :::| 
  2154153CTCCAGCGGTCAGCCCGCATACACCGGCCCCTGCCCGCCGGCGGGCACCG 2154104

     124 lyPheHisArgTyrGlnPhePheValTyr 133
          ||   |||:::|||:::|||   :::||| 
  2154103GGACACACCACTACCGGTTTACCCTCTAC 2154075

seq_name: /cgn1_7/ptodata/1/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


alignment_scores:
           Quality:  105.00            Length:     93
             Ratio:  2.283               Gaps:      6
 Percent Similarity:  49.462   Percent Identity: 34.409

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-103-840A-1/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-1  from: 1  to: 4411529

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         :::|||      |||         |||   ||||||
  2157045ACCTGGTCGGCGCCG.........TTTGGCGGCGCA.............. 2157019

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
            ||||||:::   ||||||||||||||         ||||||
  2157018.......CTCGTTGTCGATGATCCGGACGCACCT......CGCGAACCT. 2156983

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLys......Gly 98
                 :::   ||||||:::|||      |||         ::: 
  2156982...........TACGTCCATTGGATCGTGATCGGGATCGCCCCTGGTGCT 2156944

      99 AlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSer........ 112
         ::::::        |||:::      |||      :::|||        
  2156943GGCAGCACCGCCGATGGTGAGACTCCCGGTGGCGGAATCAGCCTGCCGAA 2156894

     113 ..............AlaTyrGlnAlaProSerProProAlaHisSerG 124
                       ||||||   :::|||   |||||||||   :::| 
  2156893CTCCAGCGGTCAGCCCGCATACACCGGCCCCTGCCCGCCGGCGGGCACCG 2156844

     124 lyPheHisArgTyrGlnPhePheValTyr 133
          ||   |||:::|||:::|||   :::||| 
  2156843GGACACACCACTACCGGTTTACCCTCTAC 2156815

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-47

seq_documentation_block:
; Sequence 47, Application US/08313185
; Patent No. 5851763
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,185
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-313-185-47


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-313-185-47    ..

 Align seg 1/1  to: US-08-313-185-47  from: 1  to: 700

      42 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 57
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      58 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 74
         :::   |||:::|||   :::   |||          ||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      74 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 90
         ||||||||||   :::                    ::::::   |||
     409 ACCCCGACGCGGTCGGC..................GGACTGTACGTGCAC 440

      91 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 100
         |||:::||||||   |||              :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     101 ......................LeuLysLysGlyLysIleGlnGlyG 109
                                :::::::|||      ||||||
```

```
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     109 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 125
                    |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     126 HisArgTyrGlnPhePheValTyr 133
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-11

seq_documentation_block:
; Sequence 11, Application US/08459499
; Patent No. 5871912
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart T.
;    APPLICANT:  Young, Douglas B.
;    APPLICANT:  Zhang, Ying
;    TITLE OF INVENTION:  Nucleic Acid Probes, Sequences, and Methods
;    TITLE OF INVENTION:  for Detecting Mycobacterium Tuberculosis Resistant to Isoni
;    TITLE OF INVENTION:  Amended)
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,499
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/875,940
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/929,206
;      FILING DATE:  27-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,655
;      FILING DATE:  11-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0110-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-459-499-11


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
```

```
US-09-215-435-225_COPY_23_227 x US-08-459-499-11   ..

Align seg 1/1  to: US-08-459-499-11  from: 1  to: 700

     42 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 57
        ||||||:::      :::      |||   |||      |||::::::
    321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

     58 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 74
         :::   |||:::|||   :::   |||         ||||||:::   |
    368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

     74 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 90
        ||||||||||   :::                    ::::::   |||
    409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

     91 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 100
        |||:::||||||   |||               :::   ||||||
    441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

    101 .......................LeuLysLysGlyLysIleGlnGlyG 109
                               ::::::|||     ||||||
    491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

    109 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 125
                  |||   :::|||   |||||||||   :::|||
    540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

    126 HisArgTyrGlnPhePheValTyr 133
        |||:::|||:::|||   :::|||
    576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/6A_COMB.seq:US-09-082-614A-47

seq_documentation_block:
; Sequence 47, Application US/09082614A
; Patent No. 6124098
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/082,614A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/313,185
;      FILING DATE:  12-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-082-614A-47


alignment_scores:
            Quality:   87.50            Length:    108
              Ratio:  1.620              Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-082-614A-47   ..

 Align seg 1/1  to: US-09-082-614A-47  from: 1  to: 700

      42 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 57
         ||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      58 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 74
          :::   |||:::|||   :::   |||         ||||||:::   |
     368 GTCGTCGCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      74 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 90
         ||||||||||   :::                    ::::::  |||
     409 ACCCCGACGCGGTCGGC.................GGACTGTACGTGCAC 440

      91 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 100
         |||:::||||||   |||               :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     101 .......................LeuLysLysGlyLysIleGlnGlyG 109
                                ::::::|||     ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     109 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 125
                   |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     126 HisArgTyrGlnPhePheValTyr 133
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-459-499-8

seq_documentation_block:
; Sequence 8, Application US/08459499
; Patent No. 5871912
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart T.
;    APPLICANT:  Young, Douglas B.
;    APPLICANT:  Zhang, Ying
;    TITLE OF INVENTION:  Nucleic Acid Probes, Sequences, and Methods
;    TITLE OF INVENTION:  for Detecting Mycobacterium Tuberculosis Resistant to Isoni
;    TITLE OF INVENTION:  Amended)
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,499
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/875,940
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/929,206
;      FILING DATE:  27-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,655
;      FILING DATE:  11-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0110-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4794 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-459-499-8


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-459-499-8   ..

 Align seg 1/1  to: US-08-459-499-8  from: 1  to: 4794

      42 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 57
         |||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      58 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 74
          :::   |||:::|||   :::   |||          |||||||:::   |
     368 GTCGTCGCCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      74 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 90
         ||||||||||   :::                     ::::::   |||
     409 ACCCCGACGCGGTCGGC.....................GGACTGTACGTGCAC 440

      91 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 100
         |||:::||||||   |||               :::   ||||||
     441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

     101 .......................LeuLysLysGlyLysIleGlnGlyG 109
                                :::::|||      ||||||
     491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA. 539

     109 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 125
                  |||   :::|||   |||||||||   :::|||
     540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

     126 HisArgTyrGlnPhePheValTyr 133
         |||:::|||:::|||   :::|||
     576 CACCACTACCGGTTTACCCTCTAC 599

 seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:US-08-313-185-45
```

```
seq_documentation_block:
; Sequence 45, Application US/08313185
; Patent No. 5851763
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,185
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4795 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-313-185-45


alignment_scores:
            Quality:   87.50              Length:    108
              Ratio:  1.620                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 31.481

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-08-313-185-45   ..

 Align seg 1/1  to: US-08-313-185-45  from: 1  to: 4795

      42 CysAsnAsnTyrArgGlnLysIleThrSerTrpMetGluProIleVal.. 57
         |||||||:::      :::      |||   |||      |||::::::
     321 TGCAATTCAGCTGCAAGGGGGCCAACG...TGGCCGCCACCGTTGACGTG 367

      58 .LysPheProGlyAlaValAspGlyAlaThrTyrIleLeuValMetValA 74
          :::   |||:::|||   :::   |||          |||||||:::   |
     368 GTCGTCGCCCCGCGGCGAGCGAACTGGCA.........CTCGTCGTCGATG 408

      74 spProAspAlaProSerArgAlaGluProArgGlnArgPheTrpArgHis 90
         ||||||||||   :::                     ::::::   |||
     409 ACCCCGACGCGGTCGGC.....................GGACTGTACGTGCAC 440

      91 TrpLeuValThrAspIle..............LysGlyAlaAsp..... 100
         |||:::||||||   |||               :::   ||||||
```

```
441 TGGATCGTGACCGGAATCGCCCCTGGCTCTGGCAGCACGGCGGATGGTCA 490

101 .......................LeuLysLysGlyLysIleGlnGlyG 109
                            :::::|||      ||||||
491 GACTCCTGCTGGTGGGCACAGCGTGCCGAATTCTGGTGGTCGGCAAGGA- 539

109 lnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPhe 125
                |||   :::|||   |||||||||   :::|||
540 .............TACTTCGGTCCATGCCCGCCGGCGGGCACCGGGACA 575

126 HisArgTyrGlnPhePheValTyr 133
    |||:::|||:::|||   :::|||
576 CACCACTACCGGTTTACCCTCTAC 599
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225_COPY_23_227 to: Pending_Patents_NA_Main:*   out_format : pfs

Date: Nov 15, 2001  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2173/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_12482 -NCPU=6 -ICPU=3 -LONGLOG
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: Pending_Patents_NA_Main:*
Database sequences: 17159718
Database length: 173266264
Search time (sec): 11750.060000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-124 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-471-276-7 + 1115.00 2069.97  5.9e-106  826  !
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-7 + 1115.00 2069.97  5.9e-106  826  !
/cgn1_7/ptodata/1/pna/US6008_COMB.seq:US-60-081-563-70 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-7 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-7 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-7 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-7 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-224-006-7 + 1115.00 2069.97  5.9e-106  826
/cgn1_7/ptodata/1/pna/US6027_COMB.seq:US-60-278-258-12888 + 1115.00 2069.11  6.6e-106  9
/cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-726-176-1502 + 1111.00 2061.99  1.6e-105  87
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-56 + 1109.00 2056.66  3.3e-105  1028
/cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-56 + 1109.00 2056.66  3.3e-105  1028
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-56 + 1109.00 2056.66  3.3e-105  1028
/cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-10894 + 1108.00 2056.02  3.5e-105  9
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-922-5157 + 1104.00 2048.81  8.9e-105  878
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-922-5157 + 1104.00 2048.81  8.9e-105  878
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-958-820-4 + 1104.00 2048.55  9.2e-105  903  !
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-390-126-4 + 1104.00 2048.55  9.2e-105  903  !
/cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-61 + 1089.00 2019.99  3.6e-103  952
/cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-61 + 1089.00 2019.99  3.6e-103  952
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-61 + 1089.00 2019.99  3.6e-103  952
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-7851 + 1085.00 2013.01  8.8e-103  90
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-931-836-60 + 1080.00 2003.80  2.9e-102  890
/cgn1_7/ptodata/1/pna/US096A_COMB.seq:US-09-617-746A-9 + 1076.00 1995.47  8.3e-102  972
/cgn1_7/ptodata/1/pna/US096B_COMB.seq:US-09-631-451A-9 + 1076.00 1995.47  8.3e-102  972
/cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-491-404-1525 + 1046.00 1940.25  9.9e-99  886
/cgn1_7/ptodata/1/pna/US099_COMB.seq:US-09-922-279-1525 + 1046.00 1940.25  9.9e-99  886
/cgn1_7/ptodata/1/pna/US6028_COMB.seq:US-60-281-593-409 +  844.00 1565.94  7.0e-78  615
/cgn1_7/ptodata/1/pna/US096C_COMB.seq:US-09-652-914-2551 +  679.00 1258.44  9.4e-61  54
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-2655 +  567.00 1050.32  3.7e-49  476
/cgn1_7/ptodata/1/pna/US089_COMB.seq:US-08-905-144-182 +  543.00 1006.03  1.1e-46  447
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-3691 +  535.00  992.27  6.3e-46  394
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-2207 +  503.00  932.09  1.4e-42  408
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163A-4626 +  495.50  916.43  1.1e-41  48
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163B-4626 +  495.50  916.43  1.1e-41  48
/cgn1_7/ptodata/1/pna/US082_COMB.seq:US-08-276-163D-4626 +  495.50  916.43  1.1e-41  48
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-840-145-4626 +  495.50  916.43  1.1e-41  484
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-4958 +  469.00  868.16  5.2e-39  423
/cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-4958 +  469.00  868.16  5.2e-39  42
/cgn1_7/ptodata/1/pna/US088_COMB.seq:US-08-880-314-4295 +  464.00  862.99  1.0e-38
/cgn1_7/ptodata/1/pna/US095B_COMB.seq:US-09-534-853-7719 +  464.00  862.99  1.0e-38
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-4300 +  461.00  854.84  2.8e-38
/cgn1_7/ptodata/1/pna/US093_COMB.seq:US-09-359-067-4523 +  456.00  845.57  9.4e-38
/cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-824-559-6494 +  456.00  843.73  1.2e-37

seq_name: /cgn1_7/ptodata/1/pna/US092_COMB.seq:US-09-215-435-124

seq_documentation_block:
; Sequence 124, Application US/09215435A
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert Aymeric,
;  APPLICANT: Bougueleret Lydie
;  TITLE OF INVENTION: Extended cDNAs
;  FILE REFERENCE: GENSET.019A
;  CURRENT APPLICATION NUMBER: US/09/215,435A
;  CURRENT FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/069,957
;  EARLIER FILING DATE: 1997-12-17
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-2-9
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-4-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-8-10
;  NUMBER OF SEQ ID NOS: 519
;  SOFTWARE: Patent.pm
; SEQ ID NO 124
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 795..800
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 814..826
US-09-215-435-124


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-215-435-124    ..

 Align seg 1/1  to: US-09-215-435-124  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

 67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

 84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

201 GluIleAlaAlaCys 205
     ||||||||||||||||
681 GAGATAGCTGCCTGC 695
```

seq_name: /cgn1_7/ptodata/1/pna/US094_COMB.seq:US-09-471-276-7

```
seq_documentation_block:
; Sequence 7, Application US/09471276
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Duclert A.
;  APPLICANT: Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.025CP1
;  CURRENT APPLICATION NUMBER: US/09/471,276
;  CURRENT FILING DATE: 1999-12-21
;  EARLIER APPLICATION NUMBER: 09/057,719
;  EARLIER FILING DATE: 1998-04-09
;  EARLIER APPLICATION NUMBER: 09/069,047
;  EARLIER FILING DATE: 1998-04-28
;  EARLIER APPLICATION NUMBER: PCT/IB99/00712
;  EARLIER FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 1622
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-09-471-276-7
```

```
alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-471-276-7   ..

 Align seg 1/1  to: US-09-471-276-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695
```

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-834-366-7

```
seq_documentation_block:
; Sequence 7, Application US/09834366
; GENERAL INFORMATION:
;  APPLICANT: Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
```

```
;  FILE REFERENCE: 81.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/834,366
;  CURRENT FILING DATE:  2001-04-13
;  PRIOR APPLICATION NUMBER: US 60/197,873
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 52153
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-09-834-366-7


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-834-366-7   ..

 Align seg 1/1  to: US-09-834-366-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6008_COMB.seq:US-60-081-563-70

seq_documentation_block:
; Sequence 70, Application US/60081563
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste Dumas Milne
;    APPLICANT:  Duclert, Aymeric
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/60/081,563
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  GENSET.027PR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  70:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  826 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Prostate
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..695
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  15..80
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 8.5
;      OTHER INFORMATION:  seq AALLLGLMMVVTG/DE
;    FEATURE:
;      NAME/KEY:  polya_signal
;      LOCATION:  795..800
;    FEATURE:
;      NAME/KEY:  poly_a
;      LOCATION:  814..826
;      IDENTIFICATION METHOD:  blastn
;    FEATURE:
;      NAME/KEY:  est
;      LOCATION:  426..664
;      IDENTIFICATION METHOD:  blastn
```

```
;       OTHER INFORMATION:  identity 99
;       OTHER INFORMATION:  region 99..337
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  327..435
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 1..109
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  716..777
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 100
;       OTHER INFORMATION:  region 391..452
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  660..720
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 334..394
;       OTHER INFORMATION:  id W37255
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  506..630
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 90
;       OTHER INFORMATION:  region 327..451
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  323..435
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 94
;       OTHER INFORMATION:  region 143..255
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  180..280
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 2..102
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  464..528
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 284..348
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  279..321
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 100
;       OTHER INFORMATION:  region 100..142
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  426..466
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 97
;       OTHER INFORMATION:  region 245..285
;       OTHER INFORMATION:  id W32197
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  358..674
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 141..457
;       OTHER INFORMATION:  id AA192427
```

```
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  670..779
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 37..146
;       OTHER INFORMATION:  id AA192427
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  485..602
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 224..341
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  601..709
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 91
;       OTHER INFORMATION:  region 115..223
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  704..779
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 46..121
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  392..436
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 97
;       OTHER INFORMATION:  region 393..437
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  429..465
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 94
;       OTHER INFORMATION:  region 363..399
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  464..493
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 334..363
;       OTHER INFORMATION:  id W33189
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  343..730
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 98
;       OTHER INFORMATION:  region 8..395
;       OTHER INFORMATION:  id HSPD04880
;     FEATURE:
;       NAME/KEY:  est
;       LOCATION:  507..663
;       IDENTIFICATION METHOD:  blastn
;       OTHER INFORMATION:  identity 96
;       OTHER INFORMATION:  region 264..420
;       OTHER INFORMATION:  id AA192426
US-60-081-563-70


alignment_scores:
            Quality: 1115.00                Length:    205
              Ratio:  5.439                  Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-081-563-70    ..
```

```
 Align seg 1/1  to: US-60-081-563-70  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6014_COMB.seq:US-60-147-499-7

seq_documentation_block:
; Sequence 7, Application US/60147499
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Jobert, S.
;  APPLICANT: Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/60/147,499
;  CURRENT FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-147-499-7


alignment_scores:
           Quality: 1115.00              Length:    205
             Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-147-499-7   ..

 Align seg 1/1  to: US-60-147-499-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695
```

```
seq_name: /cgn1_7/ptodata/1/pna/US6016_COMB.seq:US-60-169-629-7

seq_documentation_block:
; Sequence 7, Application US/60169629
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: GENSET.071PRF
;  CURRENT APPLICATION NUMBER: US/60/169,629
;  CURRENT FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-169-629-7


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-169-629-7   ..

 Align seq 1/1  to: US-60-169-629-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6018_COMB.seq:US-60-187-470-7

seq_documentation_block:
; Sequence 7, Application US/60187470
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Bougueleret, L.
;  APPLICANT: Jobert, S.
;  TITLE OF INVENTION: cDNAS for Secreted Proteins
;  FILE REFERENCE: 78.US2.PRO
;  CURRENT APPLICATION NUMBER: US/60/187,470
;  CURRENT FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 715
;  SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-187-470-7


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-187-470-7   ..

 Align seq 1/1  to: US-60-187-470-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

   67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

   84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

  101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

  117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

  134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

  151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

  167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

  184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

  201 GluIleAlaAlaCys 205
      ||||||||||||||||
  681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6019_COMB.seq:US-60-197-873-7

seq_documentation_block:
; Sequence 7, Application US/60197873
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  APPLICANT:  Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Jobert, Severin
;  APPLICANT:  Giordano, Jean-Yves
;  TITLE OF INVENTION:  ESTs and Encoded Human Proteins.
;  FILE REFERENCE:  81.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/197,873
;  CURRENT FILING DATE:  2000-04-18
;  NUMBER OF SEQ ID NOS:  52153
;  SOFTWARE:  Patent.pm
; SEQ ID NO 7
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-197-873-7


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0

 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-197-873-7    ..

 Align seg 1/1  to: US-60-197-873-7  from: 1  to: 826

        1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

       17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

       34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

       51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

       67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

       84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

      101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

      117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

      134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

      151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

      167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

      184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

      201 GluIleAlaAlaCys 205
          ||||||||||||||||
      681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6022_COMB.seq:US-60-224-006-7

seq_documentation_block:
; Sequence 7, Application US/60224006
; GENERAL INFORMATION:
;  APPLICANT:  Bejanin, Stephane
;  APPLICANT:  Tanaka, Hiroaki
;  TITLE OF INVENTION:  ESTs encoding human proteins
;  FILE REFERENCE:  95.US1.PRO
;  CURRENT APPLICATION NUMBER: US/60/224,006
;  CURRENT FILING DATE:  2000-08-04
;  NUMBER OF SEQ ID NOS:  67
;  SOFTWARE:  Patent.pm
; SEQ ID NO 7
```

```
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 15..695
;   NAME/KEY: sig_peptide
;   LOCATION: 15..80
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 8.5
;   OTHER INFORMATION: seq AALLLGLMMVVTG/DE
US-60-224-006-7


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-224-006-7    ..

 Align seg 1/1  to: US-60-224-006-7  from: 1  to: 826

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCCTGGACGAGGA 130

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 180

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 230

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 280

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 330

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 380

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 430

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 480

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 TTCAGGAAGGAAAGGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 530

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 580

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 630

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     631 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 680

     201 GluIleAlaAlaCys 205
```

```
         ||||||||||||||||
     681 GAGATAGCTGCCTGC 695

seq_name: /cgn1_7/ptodata/1/pna/US6027_COMB.seq:US-60-278-258-12888

seq_documentation_block:
; Sequence 12888, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms Identified Thereby
;  FILE REFERENCE: GX-0010-1 P
;  CURRENT APPLICATION NUMBER: US/60/278,258
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 17730
;  SOFTWARE: PERL Program
; SEQ ID NO 12888
;   LENGTH: 904
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 204392.4
US-60-278-258-12888


alignment_scores:
            Quality: 1115.00              Length:    205
              Ratio:  5.439                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-278-258-12888    ..

 Align seg 1/1  to: US-60-278-258-12888  from: 1  to: 904

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 204

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 254

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 304

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 354

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     355 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 404

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     405 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 454

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     455 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 504

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     505 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 554

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     555 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 604

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     605 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 654

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     655 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 704

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     705 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 754

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     755 GAGATAGCTGCCTGC 769

seq_name: /cgn1_7/ptodata/1/pna/US097B_COMB.seq:US-09-726-176-1502

seq_documentation_block:
; Sequence 1502, Application US/09726176
; GENERAL INFORMATION:
;  APPLICANT:  Gearing, David P.
;  APPLICANT:  Fraser, Christopher C.
;  APPLICANT:  Donovan, Michael J.
;  TITLE OF INVENTION:  NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE:  1600.2018-001
;  CURRENT APPLICATION NUMBER: US/09/726,176
;  CURRENT FILING DATE:  2000-11-29
;  PRIOR APPLICATION NUMBER:  60/167,841
;  PRIOR FILING DATE:  1999-11-29
;  NUMBER OF SEQ ID NOS:  1882
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1502
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-726-176-1502


alignment_scores:
            Quality: 1111.00            Length:    205
              Ratio:  5.420              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-726-176-1502   ..

 Align seg 1/1  to: US-09-726-176-1502  from: 1  to: 870

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     137 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 186

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     187 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 236

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     237 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 286

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     287 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 336

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     337 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 386

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     387 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 436

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||:::||||||||||||||||||||||||||||||||||||||||||
     437 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 486

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     487 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 536

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     537 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 586

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     587 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 636

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     637 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 686

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     687 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 736

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     737 GAGATAGCTGCCTGC 751

seq_name: /cgn1_7/ptodata/1/pna/PCTUS_COMB.seq:PCT-US99-29950-56

seq_documentation_block:
; Sequence 56, Application PC/TUS9929950
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/29950
;  CURRENT FILING DATE: 1999-12-16
;  EARLIER APPLICATION NUMBER: 60/112,809
;  EARLIER FILING DATE: 1998-12-17
;  EARLIER APPLICATION NUMBER: 60/113,006
;  EARLIER FILING DATE: 1998-12-18
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1026)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1027)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1028)
;   OTHER INFORMATION: n equals a,t,g, or c
PCT-US99-29950-56


alignment_scores:
            Quality: 1109.00            Length:    205
              Ratio:  5.436              Gaps:      0
```

```
 Percent Similarity:  99.512   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x PCT-US99-29950-56   ..

 Align seq 1/1  to: PCT-US99-29950-56  from: 1  to: 1028

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     219 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 268

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     269 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 318

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     319 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 368

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     369 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 418

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     419 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 468

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     469 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 518

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     519 CTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 568

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     569 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 618

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     619 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 668

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     669 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 718

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     719 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 768

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     769 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 818

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     819 GAGATAGCTGCCTGC 833

 seq_name: /cgn1_7/ptodata/1/pna/US095D_COMB.seq:US-09-591-316-56

 seq_documentation_block:
 ; Sequence 56, Application US/09591316
 ; GENERAL INFORMATION:
 ;  APPLICANT: Rosen et al.
 ;  TITLE OF INVENTION: 47 Human Secreted Proteins
 ;  FILE REFERENCE: PZ035P1
 ;  CURRENT APPLICATION NUMBER: US/09/591,316
 ;  CURRENT FILING DATE: 2000-06-09
 ;  EARLIER APPLICATION NUMBER: PCT/US99/29950
 ;  EARLIER FILING DATE: 1999-12-16
 ;  EARLIER APPLICATION NUMBER: 60/113,006
 ;  EARLIER FILING DATE: 1998-12-18
 ;  EARLIER APPLICATION NUMBER: 60/112,809
 ;  EARLIER FILING DATE: 1998-12-17
 ;  NUMBER OF SEQ ID NOS: 231
 ;  SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 56
 ;   LENGTH: 1028
 ;   TYPE: DNA
 ;   ORGANISM: Homo sapiens
 ;   FEATURE:
 ;   NAME/KEY: SITE
 ;   LOCATION: (1022)
 ;   OTHER INFORMATION: n equals a,t,g, or c
 ;   FEATURE:
 ;   NAME/KEY: SITE
 ;   LOCATION: (1026)
 ;   OTHER INFORMATION: n equals a,t,g, or c
 ;   FEATURE:
 ;   NAME/KEY: SITE
 ;   LOCATION: (1027)
 ;   OTHER INFORMATION: n equals a,t,g, or c
 ;   FEATURE:
 ;   NAME/KEY: SITE
 ;   LOCATION: (1028)
 ;   OTHER INFORMATION: n equals a,t,g, or c
 US-09-591-316-56


 alignment_scores:
             Quality: 1109.00              Length:    205
               Ratio:  5.436                Gaps:      0
  Percent Similarity:  99.512   Percent Identity: 99.512

 alignment_block:
  US-09-215-435-225_COPY_23_227 x US-09-591-316-56   ..

  Align seq 1/1  to: US-09-591-316-56  from: 1  to: 1028

        1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      219 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 268

       17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      269 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 318

       34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      319 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 368

       51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      369 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 418

       67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      419 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 468

       84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      469 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 518

      101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
          ||||||   ||||||||||||||||||||||||||||||||||||||||||
      519 CTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 568

      117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      569 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 618

      134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      619 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 668
```

```
     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     669 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 718

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     719 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 768

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     769 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 818

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     819 GAGATAGCTGCCTGC 833

seq_name: /cgn1_7/ptodata/1/pna/US098_COMB.seq:US-09-895-298-56

seq_documentation_block:
; Sequence 56, Application US/09895298
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 47 Human Secreted Proteins
;  FILE REFERENCE: PZ035P1
;  CURRENT APPLICATION NUMBER: US/09/895,298
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: 09/591,16
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: PCT/US99/29950
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: 60/113,006
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 60/112,809
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1028
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1022)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1026)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1027)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1028)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-895-298-56


alignment_scores:
            Quality: 1109.00             Length:    205
              Ratio:  5.436               Gaps:      0
 Percent Similarity:  99.512   Percent Identity: 99.512

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-895-298-56   ..

 Align seg 1/1  to: US-09-895-298-56  from: 1  to: 1028

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     219 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 268

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     269 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 318

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     319 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 368

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     369 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 418

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     419 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 468

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     469 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 518

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     519 CTGAAGRAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 568

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     569 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 618

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     619 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 668

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     669 GGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACC 718

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     719 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 768

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     769 TCCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 818

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     819 GAGATAGCTGCCTGC 833

seq_name: /cgn1_7/ptodata/1/pna/US6017_COMB.seq:US-60-172-373-10894

seq_documentation_block:
; Sequence 10894, Application US/60172373
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT: Lal, Preeti
;  APPLICANT: Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym
;  FILE REFERENCE: GX-0006 P
;  CURRENT APPLICATION NUMBER: US/60/172,373
;  CURRENT FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 25,772
;  SOFTWARE: PERL Program
; SEQ ID NO 10894
;   LENGTH: 904
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 204392.1
US-60-172-373-10894


alignment_scores:
            Quality: 1108.00             Length:    205
              Ratio:  5.431               Gaps:      0
```

Percent Similarity: 99.512 Percent Identity: 99.512

alignment_block:
US-09-215-435-225_COPY_23_227 x US-60-172-373-10894 ..

Align seg 1/1 to: US-60-172-373-10894 from: 1 to: 904

```
       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 204

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 254

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 304

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 354

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     355 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 404

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     405 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 454

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     455 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 504

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     505 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 554

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     555 TTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGA 604

     151 GlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     605 GGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTGGGCGAACC 654

     167 oGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     655 TGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCC 704

     184 euGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAla 200
         |||||||||||||   ||||||||||||||||||||||||||||||||||
     705 TCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCG 754

     201 GluIleAlaAlaCys 205
         ||||||||||||||||
     755 GAGATAGCTGCCTGC 769
```

```
OM of: US-09-215-435-225_COPY_23_227 to: Pending_Patents_NA_New:*   out_format : pfs

Date: Nov 15, 2001  4:11 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141105_2194/app_query.fasta_1.1519
 -DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09215435_@CGN1_1_69 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: Pending_Patents_NA_New:*
Database sequences: 127876
Database length: 122380375
Search time (sec): 564.420000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-253-167 +  205.00  439.49  2.9e-17   1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-1 +  164.00  350.79  2.5e-12   856
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-3 -  164.00  350.79  2.5e-12   856
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4131 +  158.00  336.41  1.6e-11
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-971 +  105.50  220.93  4.3e-05   5
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1413 +  103.00  221.23  4.2e-05
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3078 +   96.50  198.20  0.0008
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4132 +   92.50  187.32  0.0032
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6698 +   80.50  156.19  0.1744
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-6710 +   79.00  161.93  0.0835
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-481 -   78.50  112.60  46.71     2
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3027 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3028 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3029 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3030 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3031 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3032 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3033 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3034 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3035 -   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3036 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3037 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3038 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3039 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3040 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3041 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3042 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3043 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3044 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3045 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3046 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3047 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3048 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3049 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3050 +   78.50  112.60  46.71
/cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3678 +   78.00  154.38  0.2198
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-330 -   77.00  117.17  25.98     1
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2244 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2245 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2246 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2247 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2248 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2249 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2250 -   77.00  117.17  25.98
/cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-2251 -   77.00  117.17  25.98

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-960-253-167

seq_documentation_block:
; Sequence 167, Application US/09960253
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Lodes, Michael J.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.556
;  CURRENT APPLICATION NUMBER: US/09/960,253
;  CURRENT FILING DATE:  2001-09-20
;  NUMBER OF SEQ ID NOS: 187
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 1444
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-960-253-167


alignment_scores:
           Quality:  205.00            Length:    151
             Ratio:  2.228               Gaps:      4
 Percent Similarity:  60.927   Percent Identity: 33.113

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-960-253-167   ..

 Align seg 1/1  to: US-09-960-253-167  from: 1  to: 1444

      30 GluLeuGlyAsnIleGlyCysLysValValProAspCysAsnAsnTyrAr 46
         ||||||||||::::::         :::::::|||        :::   ||
     216 GAGCTGGGCAAAGTG.........CTGACGCCCACCCAGGTTAAGAATAG 256

      46 gGlnLysIleThrSerTrpMetGluProIleValLysPheProGlyAlaV 63
         |            ||||||                          |||
     257 ACCCACCAGCATTTCGTGG.....................GATGGTCTTG 285

      63 alAspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSer 79
           :::|||      |||   ||||||::::::||||||||||||||||||
     286 ATTCAGGGAAGCTCTACACCTTGGTCCTGACAGACCCGGATGCTCCCAGC 335

      80 ArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIl 96
         |||   :::|||:::   |||   |||:::|||:::||||||:::::::
     336 AGGAAGGATCCCAAATACAGAGAATGGCATCATTTCCTGGTGGTCAACAT 385

      96 eLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerA 113
         :||||||   |||:::::::::|||   :::            ||||||
     386 GAAGGGCAATGACATCAGCAGTGGCACAGTC............CTCTCCG 423

     113 laTyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGln 129
           |||   :::   :::||||||      :::|||:::|||||||||
     424 ATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCTCCACCGCTATGTC 473

     130 PhePheValTyrLeuGlnGluGlyLysValIleSer............Le 142
         ::::::||||||   |||:::      :::                  ||
     474 TGGCTGGTTTACGAGCAGGACAGGCCGCTAAAGTGTGACGAGCCCATCCT 523

     142 uLeuProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheL 159
         |      :::::::::      ||||||::::::|||:::     |||
     524 CAGCAACCGATCTGGAGACCACCGTGGCAAATTCAAGGTGGCGTCCTTCC 573

     159 euAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThr 175
           ::::::::::::|||       |||   |||:::|||   :::::::::
     574 GTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCACGTGTTACCAGGCC 623
```

```
     176 Gln 176
         :::
     624 GAG 626

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-1

seq_documentation_block:
; Sequence 1, Application US/09845849
; GENERAL INFORMATION:
;  APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  WEIGEL, Detlef
;  APPLICANT:  KARDAILSKY, Igor
;  TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE: SALKINS.026DV1
;  CURRENT APPLICATION NUMBER: US/09/845,849
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/060,726
;  PRIOR FILING DATE: 1998-04-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-845-849-1


alignment_scores:
            Quality:  164.00            Length:    113
              Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-845-849-1   ..

 Align seg 1/1  to: US-09-845-849-1  from: 1  to: 856

      68 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 84
         |||   |||||||||||||||||||||:::||||||   :::::::|||:
     255 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 304

      84 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 101
         :   |||   :::   ||||||||||||||||||||   ::::::
     305 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 350

     101 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 117
               |||         |||:::|||:::   :::|||:::   |||
     351 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 395

     118 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 134
         ||||||   |||      |||:::||||||      |||:::::::::
     396 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 439

     134 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 151
          |||   |||:::              |||
     440 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 470

     151 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 163
           :::|||:::::::::            :::|||   :::   ::::::
     471 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 518

     164 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 176
         ||||||   |||   |||:::::   :::      |||
     519 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 557

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-845-849-3

seq_documentation_block:
; Sequence 3, Application US/09845849
; GENERAL INFORMATION:
;  APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  WEIGEL, Detlef
;  APPLICANT:  KARDAILSKY, Igor
;  TITLE OF INVENTION: FLOWERING LOCUS T (FT) AND GENETICALLY
;  TITLE OF INVENTION:  MODIFIED PLANTS HAVING MODULATED FLOWER DEVELOPMENT
;  FILE REFERENCE: SALKINS.026DV1
;  CURRENT APPLICATION NUMBER: US/09/845,849
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 09/060,726
;  PRIOR FILING DATE: 1998-04-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 856
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-845-849-3


alignment_scores:
            Quality:  164.00            Length:    113
              Ratio:  2.412               Gaps:      4
 Percent Similarity:  60.177   Percent Identity: 36.283

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-845-849-3/rev   ..

 Align seg 1/1  to reverse of: US-09-845-849-3  from: 1  to: 856

      68 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 84
         |||   |||||||||||||||||||||:::||||||   :::::::|||:
     602 TATACTTTGGTTATGGTGGATCCAGATGTTCCAAGTCCTAGCAACCCTCA 553

      84 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 101
         :   |||   :::   ||||||||||||||||||||   ::::::
     552 CCTCCGAGAATATCTCCATTGGTTGGTGACTGATATCCCTGCTACA.... 507

     101 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 117
               |||         |||:::|||:::   :::|||:::   |||
     506 .....ACTGGAACAACCTTTGGCAATGAGATTGTGTGTTACGAAAATCCA 462

     118 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 134
         ||||||   |||      |||:::||||||      |||:::::::::
     461 AGTCCCACTGCA......GGAATTCATCGTGTCGTGTTTATATTGTTTCG 418

     134 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 151
          |||   |||:::              |||
     417 ACAGCTTGGCAGGCAAACAGTGTATGCACCA..................... 387

     151 lySerTrpLysMetAsp............ArgPheLeuAsnArgPheHis 163
           :::|||:::::::::            :::|||   :::   ::::::
     386 ..GGGTGGCGCCAGAACTTCAACACTCGCGAGTTTGCTGAGATCTACAAT 339

     164 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 176
         ||||||   |||   |||:::::   :::      |||
     338 CTCGGCCTTCCCGTGGCCGCAGTTTTCTACAATTGTCAG 300

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4131

seq_documentation_block:
; Sequence 4131, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 4131
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-4131
```

```
alignment_scores:
            Quality:  158.00              Length:    141
              Ratio:  1.859                Gaps:      4
 Percent Similarity:  60.284   Percent Identity: 29.078

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-325-448-4131   ..

 Align seg 1/1  to: US-60-325-448-4131  from: 1  to: 891

      54 GluProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLe 70
         :::|||   ::::::      |||:::          :::   :::   ||
     255 AAGCCACGCATTGAAGTCCAGGGTGGGGACATGAGATCCTTCTTCACTCT 304

      70 uValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgP 87
         |||||||:::|||||||||:::|||:::   :::::::|||      |||
     305 GGTCATGACGGACCCAGATGTGCCAGGGCCTAGTGATCCATACCTTAGGG 354

      87 heTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLys 103
                |||||||:::|||:::||||||   |||:::
     355 AGCATCTCCACTGGATCGTCAGTGATATTCCTGGCACC.........ACA 395

     104 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 120
                    |||:::|||:::   :::|||:::::|||:::|||
     396 GATGCTTCTTTTGGAAGGGAGGTGGTGAGCTACGAGAGCCCCAAGCCC.. 443

     120 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 137
            :::   |||:::||||||:::   |||   ::::::   |||:::
     444 ....AACATTGGCATTCACAGGTTCACCTTCGTGCTGTTCCAACAGAAGA 489

     137 lyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrp 153
           :::              |||          :::||||||       :::
     490 AGCGGCAGGCCATGAACCCTCCC.........TCCACCAGGGACTACTTC 530

     154 LysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSe 170
         :::      ||||||   |||:::      ||||||   |||   |||::
     531 AACACGCGCCGCTTCGCCAACGAGAACGACCTTGGCCTCCCGGTGGCTGC 580

     170 rThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaP 187
         ::::   |||   :::|||      :::::::::   :::   :::
     581 CGTCTACTTCAACGCGCAG......CGGGAGACAGCGGCACGCCGCCGCT 624

     187 roArgGluArgAlaSerGluPro 194
           |||:::::::::::::::|||
     625 GACGGAAACAATCAGCCCAACCC 647

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-971

seq_documentation_block:
; Sequence 971, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 971
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-971


alignment_scores:
            Quality:  105.50              Length:    113
              Ratio:  1.598                Gaps:      2
 Percent Similarity:  58.407   Percent Identity: 26.549

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-325-448-971   ..

 Align seg 1/1  to: US-60-325-448-971  from: 1  to: 535

      64 AspGlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerAr 80
         |||   :::::::::      ||||||:::|||||||||:::|||:::
     185 GATCTTTCTTCACATTGGTGGGTTATGACAGACCCAGATGTGCCAGGACC 234

      80 gAlaGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleL 97
          ::::::|||      |||       |||   :::   :::   :::
     235 AAGTGATCCTTATCTAAGGGAGCACCTTCATTGCATTTACGCAGGATTGT 284

      97 ysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAla 113
                    |||   :::      :::      :::|||:::   :::
     285 TACTGATATACCTGGGACAACGGATGCTTCTTTTGCGCGAGGTCATAAGC 334

     114 TyrGlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPh 130
         |||::::::|||:::|||      :::   ||:::|||||||:::   ||
     335 TATGAGAGTCCAAAGCCG......AACATTGGCATCCATAGGTTCATTTT 378

     130 ePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluA 147
         |   ::::::   |||:::   :::         :::::::|||
     379 TGTGCTCTTCAAGCAGAAGCGCAGGCAAACTGTAATTGTGCCA....... 421

     147 snLysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHis 163
           :::   |||      ::::::      ||||||   ::::::
     422 ..TCCTTCAGGGACCATTTCAACACCCGCCGGTTCGCCGAGGAGAATGAT 469

     164 LeuGlyGluProGluAlaSerThrGlnPheMetThrGln 176
         ||||||   |||   |||::::::   |||   :::|||
     470 CTTGGCCTTCCTGTGGCTGCTGTCTACTTCAATGCCCAG 508

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-1413

seq_documentation_block:
; Sequence 1413, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 1413
;   LENGTH: 321
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-60-325-448-1413


alignment_scores:
            Quality:  103.00              Length:    107
              Ratio:  1.537                Gaps:      4
 Percent Similarity:  62.617   Percent Identity: 30.841

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-325-448-1413   ..

 Align seg 1/1  to: US-60-325-448-1413  from: 1  to: 321

      71 ValMetValAspProAspAlaProSerArgAlaGluProArgGlnArgPh 87
         ||||||:::|||||||||:::|||:::   :::::::|||      |||
       1 GTTATGACAGACCCAGATGTGCCAGGACCAAGTGATCCATATCTAAGGGA 50

      87 eTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp.LeuLysLys 103
                |||   :::|||||||||||||   |||:::    :::
      51 ACACCTACATTG.ATTGTAACTGATATACCTGGAACAACGGATG...... 93

     104 GlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProPr 120
               :::   |||:::|||:::   :::|||:::::::|||:::|||
      94 ...CCTCTTTTGGGACGGGAAATCATAAGCTATGAGAGCCCAAAGCCCAG 140
```

```
    120 oAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluG 137
    141 CATT......GGTATCCACAGGTTCGTTTTTGTGCTCTTCAAGCAGAAGC 184

    137 lyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrp 153
    185 GTAGGCAGGCTGTAGTTGTGCCA.........TCCTCTAGGGATCATTTC 225

    154 LysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSe 170
    226 AATACACGCCAGTTTGCTGAGGAGAACGAACTTGGCCTTCCTGTCGCTGC 275

    170 rThrGlnPheMetThrGln 176
    276 TGTCTACTTCAATGCTCAG 294
```

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-3078

```
seq_documentation_block:
; Sequence 3078, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 3078
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-3078


alignment_scores:
            Quality:   96.50              Length:    111
              Ratio:  1.636                Gaps:      4
 Percent Similarity:  53.153   Percent Identity: 27.027

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-325-448-3078   ..

 Align seg 1/1  to: US-60-325-448-3078  from: 1  to: 627

      90 HisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIl 106
       6 CACTGGATTGTTACTGATATACCTGGGACA.........ACAGATGCTTC 46

     106 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisS 123
      47 TTTTGGACGGGAAGTCATAAGCTATGAGAGCCCAAAGCCC......AACA 90

     123 erGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGlyLysVal 139
      91 TAGGCATCCACAGGTTCATTTTTGTGCTCTTCAAGCAGAAGCGAAGGCAG 140

     140 IleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLysMetAs 156
     141 ACTGTAACTGTGCCT.........TCCTTCAGGGATCATTTCAACACCCG 181

     156 pArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerThrGlnP 173
     182 TCAGTTTGCTGCGGAGAATGACCTCGGCCTCCCTGTGGCAGCTGTCTACT 231

     173 heMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAlaProArgGlu 189
     232 TCAACTGTCAG.............................AGAGAG 248

     190 ArgAlaSerGluProLysHisLysAsnGlnAla 200
     249 ACTGCTGCCAGGAGGCGCTGAAAATCCAGCTCT 281
```

seq_name: /cgn1_7/ptodata/1/pna/US60_NEW_COMB.seq:US-60-325-448-4132

```
seq_documentation_block:
; Sequence 4132, Application US/60325448
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.R.
;  APPLICANT:  Moughamer, T.G.
;  TITLE OF INVENTION: Constitutive rice promoters
;  FILE REFERENCE: 1360.026PRV
;  CURRENT APPLICATION NUMBER: US/60/325,448
;  CURRENT FILING DATE:  2001-09-26
;  NUMBER OF SEQ ID NOS: 4708
; SEQ ID NO 4132
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-60-325-448-4132


alignment_scores:
            Quality:   92.50              Length:    159
              Ratio:  1.201                Gaps:      8
 Percent Similarity:  48.428   Percent Identity: 25.786

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-60-325-448-4132   ..

 Align seg 1/1  to: US-60-325-448-4132  from: 1  to: 718

      71 ValMetValAspProAspAlaProSerArg...AlaGluPro.ArgGlnA 86
      28 GTTCATGCGGACCCAGATGTGCCAGGGCCCTAGGTGAATCCATACCCTTA 77

      86 rgPheTrpArgHisTrpLeuVal.ThrAspIleLysGly...AlaAspLe 101
      78 GGGAGCATCTCCACTGGATCGTCTAGTGATATTCCTGGCACCACAGATGC 127

     101 uLysLysGlyLysIle..................................... 106
     128 TTCTTTTGGTAAGTCCTTTTCATGCATGACTGCTTTATCTTTTGCTGATT 177

     107 .....................................GlnGlyGln 109
     178 GGTCCCATCCTTAATTTCCCTGAATAAACACATGAATGCATGCAGGAAGG 227

     110 GluLeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHi 126
     228 GAGGTGGTGAGCTACGAGAGCCCCAAGCCC......AACATTGGCATTCA 271

     126 sArgTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuL 143
     272 CAGGTTCACCTTCGTGCTGTTCCAACAGAAGAAGCGGCAGGCCATGAACC 321

     143 euProLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeu 159
     322 CTCCC.........TCCACCAGGGACTACTTCAACACACGCCGCTTCGCT 362

     160 AsnArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGl 176
     363 GACGAGAACGACCTTGGCCTCCCGGTCGCTGCCGTCTACTTC........ 404

     176 nAsnTyrGlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerG 193
     405 .......................AACGCGCAGCGGGAGACAGCGGCAC 429

     193 luProLysHisLysAsnGlnAla 200
     430 GCCGCCGCTGACGGAAACAATCA 452

 seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-815-242-6698

 seq_documentation_block:
 ; Sequence 6698, Application US/09815242
```

```
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6698
;   LENGTH: 951
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(951)
US-09-815-242-6698


alignment_scores:
           Quality:   80.50            Length:    218
             Ratio:  0.782              Gaps:      9
 Percent Similarity:  47.248   Percent Identity: 20.642

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-815-242-6698   ..

 Align seg 1/1  to: US-09-815-242-6698  from: 1  to: 951

      11 GluAlaLeuLeuAspGluAspThrLeuPheCysGlnGlyLeuGluValPh 27
         ||||||||||||||||||||||||||:::      ::::::   :::   ::
     358 GAAGCCTTACTAGACGAAGACACTGTTGTAATCGATGCTCGTAACGATTA 407

      27 eTyrProGluLeuGlyAsnIleGlyCysLysValValProAspCysAsnA 44
         :      :::||||||::::::         |||   ||||||   ::::
     408 TGAATATGATTTAGGTCATTTCCGTGGTGCCGTGCGCCCAGATATCCGTA 457

      44 snTyrArgGlnLysIleThrSerTrpMetGluProIleValLysPhePro 60
         :::::|||:::   :::   :::|||::::::
     458 GCTTCCGTGAG...TTACCACAATGGATTCGC..................... 486

      61 GlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAspAl 77

     486 ............................................... 486

      77 aProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValT 94
                    |||   ::::::::::|||           :::||||
     487 .............GAGAACAAAGAACAATTTATGGATAAAAAAATTGTTA 523

      94 hrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu 110
         ||       |||:::   ::::::      |||::::::|||
     524 CCTATTGTACTGGCGGG...ATTCGCTGTGAAAAATTTTCTGGCTGGTTA 570

     111 LeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisAr 127
         |||:::                          |||:::|||:::   ::
     571 TTAAAAGAAGGATTTGAAGATGTTGCTCAATTGCATGGTGGTATCGCCAA 620

     127 gTyr............GlnPhePheValTyrLeuGlnGluGlyLysVal. 139
         :|||          :::          |||   :::||||||:::
     621 CTATGGGAAAAATCCAGAAACACGTGGCGAACTTTGGGACGGCAAAATGT 670

     140 ......................IleSerLeuLeuProLysGluAsnLys 148
                                  :::::::::         ::::::|||
     671 ATGTCTTTGATGACCGAATCAGTGTCGAAATTAATCATGTTGATAAAAAA 720

     149 ThrArgGly...SerTrp............LysMetAspArgPheLeuAs 160
         :::   |||   :::|||                  :::|||::::::||
     721 GTTATTGGGAAAGACTGGTTTGATGGAACACCTTGCGAACGCTACATTAA 770

     160 nArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnA 177
         |            ::::::||||||::::::   |||:::::::|||::::
     771 C.........TGTGCAAACCCAGAATGTAATCGTCAAATCTTAACTTCAG 811

     177 snTyrGlnAspSerProThrLeu.......................Gln 185
         ::   ::::::        |||                         |||
     812 AAGAAAATGAACATAAACATTTAGGTGGCTGCTCATTAGAATGTAGCCAG 861

     186 AlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaGluIl 202
            |||   :::|||      :::      :::   ::::::||||||::
     862 CATCCTGCCAACCGTTATGTAAAAAAATATAATTTAACAGAAGCAGAAGT 911

     202 eAla 203
         :|||
     912 TGCT 915

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-922-340-6710

seq_documentation_block:
; Sequence 6710, Application US/09922340
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
;  TITLE OF INVENTION:  From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain
;  FILE REFERENCE: 20411-726CON2
;  CURRENT APPLICATION NUMBER: US/09/922,340
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: 09/353,690
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 09/004,182
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 12181
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6710
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-340-6710


alignment_scores:
           Quality:   79.00            Length:     47
             Ratio:  3.038              Gaps:      1
 Percent Similarity:  55.319   Percent Identity: 36.170

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-922-340-6710   ..

 Align seg 1/1  to: US-09-922-340-6710  from: 1  to: 439

      90 HisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGlyLysIl 106
         |||||||||:::|||:::|||   |||      :::
      12 CACTGGCTGCTAACCAACATCCCGGGTAACCGGGTG............GC 49

     106 eGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProAlaHisS 123
          :::||||||            |||      |||   |||         |
```

```
     50 TGAAGGACAGGTGACGTGTCCCTACCTCCCCCCCTTCCCTGCCCGAGGCT 99

    123 erGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGlu 136
        |||||:::||||||      |||:::::::::   |||:::
    100 CCGGCATCCACCGTCTTGCCTTCCTGCTCTTCAAGCAGGAC 140

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-481

seq_documentation_block:
; Sequence 481, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 481
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20774)..(21481)
US-09-897-516-481


alignment_scores:
            Quality:   78.50            Length:    130
              Ratio:  1.246              Gaps:      5
 Percent Similarity:  48.462   Percent Identity: 22.308

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-897-516-481/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-481  from: 1  to: 24961

      67 ThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
         ::::::   :::::::::   ||||||||||||||||:::   :::
   22653 AGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAACGGGCAGT..... 22609

      83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIle..... 96
                 ||||||   ||||||::::::   ||||||
   22608 .......GGTTTCTGG...CATTGGGTGGCATTTGATATTCCGCTGGAAA 22569

      97 ..............................LysGlyAlaAspLeuLys 102
                                        |||:::   |||
   22568 TACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGGCAGCAAATTACCG 22519

     103 LysGlyLysIleGlnGlyGlnGlu.........LeuSerAlaTyrGlnAl 116
           |||   ||||||:::::::::               ::::::   ::
   22518 GCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGTTTGGTTTCGGTGG 22469

     116 aProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValT 133
         :      ||||||      :::      |||||||||   |||   |||:
   22468 AGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTATATTTTCACGGTAC 22419

     133 yrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         ::   ::::::   :::   :::::::::
   22418 ATGCCATGTCGGTAGAAAAACTGGGTATT..................... 22390

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
               :::   ::::::   :::   ::::::||||||   :::
   22389 ......AACGCTGAAATAACAAATGCGGTCACACGTTTTATGATCTACGC 22346

     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGln 179
            :::   :::|||      :::|||      ||||||
   22345 CAATAGTCTGGCAACCGCCACTATTACAGGCTATTACCAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3027

seq_documentation_block:
; Sequence 3027, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3027
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (338)..(1009)
US-09-897-516-3027


alignment_scores:
            Quality:   78.50            Length:    130
              Ratio:  1.246              Gaps:      5
 Percent Similarity:  48.462   Percent Identity: 22.308

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-897-516-3027/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3027  from: 1  to: 24961

      67 ThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
         ::::::   :::::::::   ||||||||||||||||:::   :::
   22653 AGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAACGGGCAGT..... 22609

      83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIle..... 96
                 ||||||   ||||||::::::   ||||||
   22608 .......GGTTTCTGG...CATTGGGTGGCATTTGATATTCCGCTGGAAA 22569

      97 ..............................LysGlyAlaAspLeuLys 102
                                        |||:::   |||
   22568 TACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGGCAGCAAATTACCG 22519

     103 LysGlyLysIleGlnGlyGlnGlu.........LeuSerAlaTyrGlnAl 116
           |||   ||||||:::::::::               ::::::   ::
   22518 GCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGTTTGGTTTCGGTGG 22469

     116 aProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValT 133
         :      ||||||      :::      |||||||||   |||   |||:
   22468 AGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTATATTTTCACGGTAC 22419

     133 yrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         ::   ::::::   :::   :::::::::
   22418 ATGCCATGTCGGTAGAAAAACTGGGTATT..................... 22390

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
               :::   ::::::   :::   ::::::||||||   :::
   22389 ......AACGCTGAAATAACAAATGCGGTCACACGTTTTATGATCTACGC 22346
```

```
     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGln 179
           :::   :::|||      :::|||      ||||||
   22345 CAATAGTCTGGCAACCGCCACTATTACAGGCTATTACCAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3028

seq_documentation_block:
; Sequence 3028, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3028
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2388)..(3563)
US-09-897-516-3028


alignment_scores:
           Quality:   78.50            Length:    130
             Ratio:  1.246              Gaps:      5
 Percent Similarity:  48.462   Percent Identity: 22.308

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-897-516-3028/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3028  from: 1  to: 24961

      67 ThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
         ::::::   :::::::::   ||||||||||||||||:::   :::
   22653 AGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAACGGGCAGT..... 22609

      83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIle..... 96
                 ||||||   ||||||::::::   ||||||
   22608 .......GGTTTCTGG...CATTGGGTGGCATTTGATATTCCGCTGGAAA 22569

      97 ............................LysGlyAlaAspLeuLys 102
                                     |||:::   |||
   22568 TACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGGCAGCAAATTACCG 22519

     103 LysGlyLysIleGlnGlyGlnGlu.........LeuSerAlaTyrGlnAl 116
           |||   ||||||:::::::::              ::::::   ::
   22518 GCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGTTTGGTTTCGGTGG 22469

     116 aProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValT 133
         :      ||||||      :::      |||||||||   |||   |||:
   22468 AGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTATATTTTCACGGTAC 22419

     133 yrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         ::   ::::::   :::   :::::::::
   22418 ATGCCATGTCGGTAGAAAAACTGGGTATT..................... 22390

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
           :::   ::::::   :::   ::::::||||||   :::
   22389 ......AACGCTGAAATAACAAATGCGGTCACACGTTTTATGATCTACGC 22346

     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGln 179
           :::   :::|||      :::|||      ||||||
   22345 CAATAGTCTGGCAACCGCCACTATTACAGGCTATTACCAG 22306

seq_name: /cgn1_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3029

seq_documentation_block:
; Sequence 3029, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3029
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3565)..(3777)
US-09-897-516-3029


alignment_scores:
           Quality:   78.50            Length:    130
             Ratio:  1.246              Gaps:      5
 Percent Similarity:  48.462   Percent Identity: 22.308

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-897-516-3029/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3029  from: 1  to: 24961

      67 ThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
         ::::::   :::::::::   ||||||||||||||||:::   :::
   22653 AGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAACGGGCAGT..... 22609

      83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIle..... 96
                 ||||||   ||||||::::::   ||||||
   22608 .......GGTTTCTGG...CATTGGGTGGCATTTGATATTCCGCTGGAAA 22569

      97 ............................LysGlyAlaAspLeuLys 102
                                     |||:::   |||
   22568 TACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGGCAGCAAATTACCG 22519

     103 LysGlyLysIleGlnGlyGlnGlu.........LeuSerAlaTyrGlnAl 116
           |||   ||||||:::::::::              ::::::   ::
   22518 GCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGTTTGGTTTCGGTGG 22469

     116 aProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValT 133
         :      ||||||      :::      |||||||||   |||   |||:
   22468 AGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTATATTTTCACGGTAC 22419

     133 yrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         ::   ::::::   :::   :::::::::
   22418 ATGCCATGTCGGTAGAAAAACTGGGTATT..................... 22390

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
           :::   ::::::   :::   ::::::||||||   :::
   22389 ......AACGCTGAAATAACAAATGCGGTCACACGTTTTATGATCTACGC 22346

     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGln 179
           :::   :::|||      :::|||      ||||||
```

```
   22345 CAATAGTCTGGCAACCGCCACTATTACAGGCTATTACCAG 22306

seq_name: /cgnl_7/ptodata/1/pna/US09_NEW_COMB.seq:US-09-897-516-3030

seq_documentation_block:
; Sequence 3030, Application US/09897516
; GENERAL INFORMATION:
;  APPLICANT: Corbin, David R.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Huesing, Joseph E.
;  APPLICANT:  Krasomil-Osterfeld, Karina C.
;  APPLICANT:  Malvar, Thomas M.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Spiridonov, Sergei
;  TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
;  FILE REFERENCE: 38-21(51847)B
;  CURRENT APPLICATION NUMBER: US/09/897,516
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215, 161
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 3030
;   LENGTH: 24961
;   TYPE: DNA
;   ORGANISM: Xenorhabdus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3774)..(5156)
US-09-897-516-3030


alignment_scores:
            Quality:   78.50              Length:    130
              Ratio:  1.246                Gaps:      5
 Percent Similarity:  48.462   Percent Identity: 22.308

alignment_block:
 US-09-215-435-225_COPY_23_227 x US-09-897-516-3030/rev   ..

 Align seg 1/1  to reverse of: US-09-897-516-3030  from: 1  to: 24961

      67 ThrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPr 83
         ::::::   :::::::::   ||||||||||||||||:::   :::
   22653 AGTTTTGCGGTCACGGTTTATGATCCTGATGCCCCAACGGGCAGT..... 22609

      83 oArgGlnArgPheTrpArgHisTrpLeuValThrAspIle..... 96
               ||||||   ||||||::::::   ||||||
   22608 .......GGTTTCTGG...CATTGGGTGGCATTTGATATTCCGCTGGAAA 22569

      97 ..............................LysGlyAlaAspLeuLys 102
                                      |||:::   |||
   22568 TACAGGCTTTATCCCCGGATGCCGGACGGAGTGATGGCAGCAAATTACCG 22519

     103 LysGlyLysIleGlnGlyGlnGlu.........LeuSerAlaTyrGlnAl 116
           |||   ||||||:::::::::              ::::::   ::
   22518 GCTGGCATCATTCAGAGCAGAAATGATTACGGGCAGTTTGGTTTCGGTGG 22469

     116 aProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValT 133
         :      ||||||      :::      |||||||||   |||   |||:
   22468 AGCCTGCCCGCCAGAGGGAGATAAAGCCCATCGCTATATTTTCACGGTAC 22419

     133 yrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         ::   ::::::   :::   :::::::::
   22418 ATGCCATGTCGGTAGAAAAACTGGGTATT..................... 22390

     150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
              :::   ::::::   :::   ::::::||||||   :::
   22389 ......AACGCTGAAATAACAAATGCGGTCACACGTTTTATGATCTACGC 22346

     166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGln 179
           :::   :::|||      :::|||      ||||||
   22345 CAATAGTCTGGCAACCGCCACTATTACAGGCTATTACCAG 22306
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-215-435-225_COPY_23_227 to: EST:*   out_format : pfs

Date: Nov 14, 2001 10:31 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US09215435/runat_14112001_141104_2118/app_query.fasta_1.1519
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09215435_@CGN1_1_6733 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-215-435-225_COPY_23_227
Query length: 205
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 5896.450000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est102:BG545668   +  919.50 1857.96  2.3e-94   749  ! BG545668 602572933F1 NIH_MGC_77
gb_est93:BF854481    +  918.00 1858.01  2.3e-94   564  ! BF854481 MR2-EN0093-271200-005-
gb_est93:BF852989    +  775.00 1565.19  4.6e-78   603  ! BF852989 MR2-EN0093-191200-001-
gb_est93:BF852995    +  744.00 1502.62  1.4e-74   563  ! BF852995 MR2-EN0093-191200-001-
gb_est93:BF853000    +  727.00 1468.00  1.2e-72   558  ! BF853000 MR2-EN0093-191200-001-
gb_est25:AI816715    -  718.00 1449.83  1.2e-71   547  ! AI816715 wj43b08.x1 NCI_CGAP_Lu
gb_est25:AI857902    -  718.00 1449.83  1.2e-71   547  ! AI857902 wj68a01.x1 NCI_CGAP_Lu
gb_est69:BE044451    -  709.00 1431.94  1.2e-70   523  ! BE044451 ho45g08.x1 Soares_NFL_
gb_est17:AI218954    -  699.00 1410.92  1.8e-69   552  ! AI218954 qg72g08.x1 Soares_NFL_
gb_est26:AI936064    -  695.00 1403.12  4.9e-69   534  ! AI936064 wo47h12.x1 NCI_CGAP_Ga
gb_est93:BF846558    +  687.50 1387.84  3.5e-68   532  ! BF846558 QV0-EN0058-261000-441-
gb_est110:W32197     +  686.00 1386.59  4.1e-68   451  ! W32197 zc04f05.r1 Soares_parath
gb_est93:BF853920    +  683.00 1379.35  1.0e-67   499  ! BF853920 MR2-EN0093-261200-004-
gb_est93:BF846063    -  681.00 1376.06  1.6e-67   464  ! BF846063 QV0-EN0055-181000-439-
gb_est93:BF853928    +  676.00 1364.27  7.2e-67   536  ! BF853928 MR2-EN0093-261200-004-
gb_est10:AA688029    -  675.00 1362.75  8.7e-67   511  ! AA688029 nv16e12.s1 NCI_CGAP_Pr
gb_est93:BF848190    +  672.00 1357.70  1.7e-66   463  ! BF848190 QV0-EN0055-031100-480-
gb_est93:BF848191    +  671.00 1355.66  2.2e-66   463  ! BF848191 QV0-EN0055-031100-480-
gb_est93:BF846059    +  650.00 1312.79  5.3e-64   462  ! BF846059 QV0-EN0055-181000-439-
gb_est93:BF846065    -  643.50 1299.39  2.9e-63   467  ! BF846065 QV0-EN0055-181000-439-
gb_est3:AA192426     +  627.00 1266.60  2.0e-61   430  ! AA192426 zq01g08.r1 Stratagene
gb_est110:W37255     +  626.00 1264.01  2.8e-61   452  ! W37255 zc09f01.r1 Soares_parath
gb_est93:BF848198    +  624.00 1259.66  4.8e-61   463  ! BF848198 QV0-EN0055-031100-480-
gb_est104:F20527     +  605.00 1222.59  5.6e-59   395  ! F20527 HSPD04880 HM3 Homo sapie
gb_est104:F28028     +  604.00 1219.80  8.0e-59   423  ! F28028 HSPD16404 HM3 Homo sapie
gb_est53:AW975331    -  582.00 1173.96  2.8e-56   459  ! AW975331 EST387439 MAGE reseque
gb_est104:F24489     +  581.00 1175.41  2.4e-56   334  ! F24489 HSPD10829 HM3 Homo sapie
gb_est101:BG483879   +  569.50 1142.49  1.6e-54   789  ! BG483879 602504455F1 NIH_MGC_77
gb_est3:AA192427     -  556.00 1120.90  2.6e-53   457  ! AA192427 zq01g08.s1 Stratagene
gb_htc:AK006964      +  536.00 1070.95  1.6e-50  1048  ! AK006964 Mus musculus adult ma
gb_est104:F33888     +  529.00 1067.16  2.5e-50   402  ! F33888 HSPD28217 HM3 Homo sapie
gb_est13:AA903899    -  491.50  989.57  5.3e-46   440  ! AA903899 oe78e07.s1 NCI_CGAP_Lu
gb_est104:F26660     +  490.00  991.47  4.2e-46   280  ! F26660 HSPD14219 HM3 Homo sapie
gb_est12:AA854779    -  487.00  981.10  1.6e-45   412  ! AA854779 aj77b05.s1 Soares_para
gb_est104:F19507     +  481.00  973.20  4.3e-45   277  ! F19507 HSPD04111 HM1 Homo sapie
gb_est104:F36903     +  479.00  964.14  1.4e-44   436  ! F36903 HSPD34824 HM3 Homo sapie
gb_est10:AA661735    -  474.00  954.87  4.6e-44   400  ! AA661735 nv04b11.s1 NCI_CGAP_Pr
gb_est90:BF599823    +  472.50  951.46  7.1e-44   413  ! BF599823 264147 MARC 3BOV Bos t
gb_est93:BF853537    +  451.00  907.70  1.9e-41   407  ! BF853537 MR2-EN0093-211200-002-
gb_est71:BE165776    +  449.00  903.78  3.2e-41   401  ! BE165776 CM3-HT0486-310100-075-
gb_est105:H83927     -  444.00  895.00  9.8e-41   352  ! H83927 ys64b02.s1 Soares retina
gb_est110:W33189     -  444.00  892.63  1.3e-40   437  ! W33189 zc09f01.s1 Soares_par
gb_est104:F31628     +  436.00  882.07  5.2e-40   258  ! F31628 HSPD22948 HM3 Homo sa
gb_est93:BF890431    +  426.00  855.44  1.6e-38   454  ! BF890431 292000 MARC 3BOV Bo
gb_est105:H83784     +  400.00  802.68  1.4e-35   440  ! H83784 ys64b02.r1 Soares ret

seq_name: gb_est102:BG545668

seq_documentation_block:
LOCUS        BG545668      749 bp    mRNA            EST       04-APR-2001
DEFINITION   602572933F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4700840 5',
             mRNA sequence.
ACCESSION    BG545668
VERSION      BG545668.1  GI:13544333
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 749)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONTECH Laboratories, Inc.
              cDNA Library Preparation: CLONTECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1535  row: b  column: 09
             High quality sequence stop: 740.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4700840"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      184 a    206 c    205 g    154 t
ORIGIN


alignment_scores:
            Quality:  919.50             Length:    202
              Ratio:  4.839                Gaps:      5
 Percent Similarity:  94.059   Percent Identity: 92.574

alignment_block:
 US-09-215-435-225_COPY_23_227 x BG545668   ..

 Align seg 1/1  to: BG545668  from: 1  to: 749

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 201

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         |||||||||||||   ||||||||||||||||||||||||||||||||||
     202 CACCCTCTTTTGCCA.GGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 250

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 300

  51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 350

  67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 400

  84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 450

 101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
     ||||||:::||||||||||||||||||||||||||||||||||||||||||
 451 CTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 500

 117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
     |||||||   ||||||||||||||||||||||||||||||||||||||||||
 501 CTCCCCAACGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 550

 134 euGlnGluGly.LysValIleSerLeuLeuProLysGluAsnLysThrAr 150
     |||||||||||  :::||||||||||||||||||||||||||||||||||
 551 TTCAGGAAGGAACAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCG 600

 150 gGlySerTrpLysMetAspArg...PheLeuAsnArgPheHisLeuGlyG 166
     ||||||||||:::           ||||||||||||||||||||||||||
 601 AGGCTCTTGGAACAATGGCACAGATTTCTTGAACCGGTTCCACCTGGGCG 650

 166 luProGluAlaSerThrGlnPheMet.ThrGlnAsn.TyrGlnAspSer. 181
     ||||||||||||||||||||||||||| |||||||||| ||||||||||||
 651 AACCTGAAGCAAGCACCCAGTTCATGAACCCAGAACTTACCAGGACTCAC 700

 182 ProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLys 197
     ||||||||||||||||||||   ||||||||||||||      |||
 701 CCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCAGCACCAAAA 748
```

seq_name: gb_est93:BF854481

seq_documentation_block:

```
LOCUS       BF854481      564 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-271200-005-d12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF854481
VERSION     BF854481.1  GI:12242225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-EN0093-
            271200-005-d12&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      149 a    156 c    141 g    118 t
ORIGIN
```

alignment_scores:

```
            Quality:  918.00              Length:    173
              Ratio:  5.368                 Gaps:      1
 Percent Similarity:  98.844   Percent Identity: 98.844
```

alignment_block:

US-09-215-435-225_COPY_23_227 x BF854481 ..

Align seg 1/1 to: BF854481 from: 1 to: 564

```
   5 AsnSerProCysAlaHisGluAlaLeuLeuAspGluAspThrLeuPheCy 21
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  46 AACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGACACCCTCTTTTG 95

  21 sGlnGlyLeuGluValPheTyrProGluLeuGlyAsnIleGlyCysLysV 38
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  96 CCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACATTGGCTGCAAGG 145

  38 alValProAspCysAsnAsnTyrArgGlnLysIleThrSerTrpMetGlu 54
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 146 TTGTTCCTGATTGTAACAACTACAGACAGAAGATCACCTCCTGGATGGAG 195

  55 ProIleValLysPheProGlyAlaValAspGlyAlaThrTyrIleLeuVa 71
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 196 CCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAACCTATATCCTGGT 245

  71 lMetValAspProAspAla.ProSerArgAlaGluProArgGlnArgPhe 87
     |||||||||||||||||| ||||||||||||||||||||||||||||||
 246 GATGGTGGATCCAGATGCCCCCTAGCAGAGCAGAACCCAGACAGAGATTC 295

  88 TrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysLysGl 104
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 296 TGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGACCTGAAGAAAGG 345

 104 yLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerProProA 121
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 346 GAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCCCTCCCCACCGG 395

 121 laHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGluGly 137
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 396 CACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCTTCAGGAAGGA 445

 138 LysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerTrpLy 154
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 446 AAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGAGGCTCTTGGAA 495

 154 sMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAlaSerT 171
     |||||||||||||||||||||||||||||||||   ||||||||||
 496 AATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAAACTGAAGCAAGCA 545

 171 hrGlnPheMetThrGln 176
     ||||||||||||||||
 546 CCCAGTTCATGACCCAG 562
```

```
seq_name: gb_est93:BF852989

seq_documentation_block:
LOCUS       BF852989      603 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-a12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852989
VERSION     BF852989.1  GI:12240837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-a12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 560.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      140 a    157 c    171 g    135 t
ORIGIN


alignment_scores:
           Quality:  775.00            Length:    149
             Ratio:  5.345               Gaps:      0
 Percent Similarity:  97.315   Percent Identity: 96.644

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF852989   ..

 Align seg 1/1  to: BF852989  from: 1  to: 603

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||:::|||||
     157 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGAACGAGGA 206

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 256

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 306

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 356

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 406

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 456

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 CTGAAGAAAGGGAAGATTCAGGGGCAGGAGTTATCAGCCTACCAGGCTCC 506

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         ||||||||||||||||||||||||||||||||   ||||||||||||||||
     507 CTCCCCACCGGCACACAGTGGCTTCCATCGCTAACAGTTCTTTGTCTATC 556

     134 euGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr 149
         |||||   ||||||||||||||||||||   |||   |||||||||
     557 TTCAGGGAGGGAAAGTCATCTCTCTCCTTTCCAAGGGAAACAAAACT 603

seq_name: gb_est93:BF852995

seq_documentation_block:
LOCUS       BF852995      563 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-e01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852995
VERSION     BF852995.1  GI:12240843
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            191200-001-e01&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
```

```
                derived from  ORESTES PCR (U.S. Letters Patent application
                No. 196,716 -  Ludwig Institute for Cancer Research)
                profiles into  the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      126 a    148 c    164 g    125 t
ORIGIN


alignment_scores:
            Quality:  744.00             Length:    135
              Ratio:  5.511                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF852995   ..

 Align seg 1/1  to: BF852995  from: 1  to: 563

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 206

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 256

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 306

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 356

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 406

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 456

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 506

     117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrL 134
         |||| |||||||||||||||||||||||||||||||||||||||||||||
     507 CTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATC 556

     134 euGln 135
         |||||
     557 TTCAG 561

seq_name: gb_est93:BF853000

seq_documentation_block:
LOCUS       BF853000      558 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-191200-001-f12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853000
VERSION     BF853000.1  GI:12240848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR2&t2=MR2-EN0093-
            191200-001-f12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 558.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      124 a    146 c    163 g    124 t      1 others
ORIGIN


alignment_scores:
            Quality:  727.00             Length:    133
              Ratio:  5.508                Gaps:      0
 Percent Similarity:  99.248   Percent Identity: 99.248

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF853000   ..

 Align seg 1/1  to: BF853000  from: 1  to: 558

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 208

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 258

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     259 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 308

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     309 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 358

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     359 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 408

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     409 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 458

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     459 CTGAAGAAAGGGAAGATTCAGGGNCAGGAGTTATCAGCCTACCAGGCTCC 508
```

```
    117 oSerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyr 133
        |||||||   ||||||||||||||||||||||||||||||||||||||
    509 CTCCCCAGCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTAT 557

seq_name: gb_est25:AI816715

seq_documentation_block:
LOCUS       AI816715      547 bp    mRNA            EST       21-DEC-1999
DEFINITION  wj43b08.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405559 3'
            similar to SW:D2_ONCVO P54187 D2 PROTEIN ;, mRNA sequence.
ACCESSION   AI816715
VERSION     AI816715.1  GI:5435794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 903   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2405559"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      102 a    131 c    160 g    154 t
ORIGIN


alignment_scores:
            Quality:  718.00            Length:    141
              Ratio:  5.129               Gaps:      0
 Percent Similarity:  99.291   Percent Identity: 98.582

alignment_block:
 US-09-215-435-225_COPY_23_227 x AI816715/rev   ..

 Align seg 1/1  to reverse of: AI816715  from: 1  to: 547

     65 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    547 GGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGC 498

     81 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    497 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 448

     98 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 114
        ||||||||||||||:::||||||||||||||||||||||||||||||||
    447 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 398

    115 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    397 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 348

    131 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    347 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 298

    148 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 164
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    297 AAACTCGAGGCTCTTGGAAAATGGACAGATTTTTGAACCGCTTCCACTTG 248

    165 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    247 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 198

    181 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 198
        ||||||||||||||||||||   ||||||||||||||||||||   |
    197 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 149

    198 snGlnAlaGluIleAlaAlaCys 205
        |||||||||||||||||||||
    148 ACCAGGCGGAGATAGCTGCCTGC 126

seq_name: gb_est25:AI857902

seq_documentation_block:
LOCUS       AI857902      547 bp    mRNA            EST       21-DEC-1999
DEFINITION  wj68a01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407944 3'
            similar to SW:D2_ONCVO P54187 D2 PROTEIN ;, mRNA sequence.
ACCESSION   AI857902
VERSION     AI857902.1  GI:5511518
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 752   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2407944"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
```

```
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
                     normalization. Library constructed by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      102 a    130 c    160 g    155 t
ORIGIN


alignment_scores:
            Quality:  718.00            Length:    141
              Ratio:  5.129               Gaps:      0
 Percent Similarity:  99.291   Percent Identity: 98.582

alignment_block:
 US-09-215-435-225_COPY_23_227 x AI857902/rev   ..

 Align seg 1/1  to reverse of: AI857902  from: 1  to: 547

      65 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     547 GGCGCAACCTATATCTTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGC 498

      81 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     497 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 448

      98 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 114
         ||||||||||||||:::||||||||||||||||||||||||||||||||
     447 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 398

     115 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     397 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 348

     131 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     347 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 298

     148 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     297 AAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTG 248

     165 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     247 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 198

     181 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 198
         ||||||||||||||||||||||   ||||||||||||||||||||||   |
     197 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 149

     198 snGlnAlaGluIleAlaAlaCys 205
         |||||||||||||||||||||||
     148 ACCAGGCGGAGATAGCTGCCTGC 126

 seq_name: gb_est69:BE044451

 seq_documentation_block:
 LOCUS       BE044451      523 bp    mRNA            EST       08-JUN-2000
 DEFINITION  ho45g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
             IMAGE:3040382 3' similar to SW:D1_ONCVO P54186 D1 PROTEIN ;, mRNA
             sequence.
 ACCESSION   BE044451
 VERSION     BE044451.1  GI:8361504
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 523)
   AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
   TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
   JOURNAL   Unpublished (1997)
 COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Seq primer: -40UP from Gibco
             High quality sequence stop: 465.
 FEATURES             Location/Qualifiers
      source          1. .523
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3040382"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo.  "
 BASE COUNT       96 a    128 c    159 g    140 t
 ORIGIN


 alignment_scores:
             Quality:  709.00            Length:    141
               Ratio:  5.101               Gaps:      0
  Percent Similarity:  98.582   Percent Identity: 97.163

 alignment_block:
  US-09-215-435-225_COPY_23_227 x BE044451/rev   ..

  Align seg 1/1  to reverse of: BE044451  from: 1  to: 523

       65 GlyAlaThrTyrIleLeuValMetValAspProAspAlaProSerArgAl 81
          ||||||||||||||||   :::||||||||||||||||||||||||||||||
      523 GGCGCAACCTATATCTGGCTGATGGTGGATCCAGATGCCCCTAGCAGAGC 474

       81 aGluProArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysG 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      473 AGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGG 424

       98 lyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyr 114
          ||||||||||||||:::||||||||||||||||||||||||||||||||
      423 GCGCCGACCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTAC 374

      115 GlnAlaProSerProProAlaHisSerGlyPheHisArgTyrGlnPhePh 131
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      373 CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTT 324

      131 eValTyrLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnL 148
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      323 TGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACA 274

      148 ysThrArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeu 164
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      273 AAACTCGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGCTTCCACCTG 224

      165 GlyGluProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSe 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      223 GGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTC 174
```

```
     181 rProThrLeuGlnAlaProArgGluArgAlaSerGluProLysHisLysA 198
         ||||||||||||||||||||||   ||||||||||||||||||||||   |
     173 ACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA.A 125

     198 snGlnAlaGluIleAlaAlaCys 205
         ||||||||||||||||||||||
     124 ACCAGGCGGAGATAGCTGCCTGC 102

seq_name: gb_est17:AI218954

seq_documentation_block:
LOCUS       AI218954      552 bp    mRNA            EST       21-DEC-1998
DEFINITION  qg72g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1840766 3' similar to SW:PBP_RAT P31044
            PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI218954
VERSION     AI218954.1  GI:3801157
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1054   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1840766"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       98 a    129 c    158 g    167 t
ORIGIN


alignment_scores:
            Quality:  699.00            Length:    138
              Ratio:  5.216               Gaps:      0
 Percent Similarity:  97.101   Percent Identity: 96.377

alignment_block:
 US-09-215-435-225_COPY_23_227 x AI218954/rev   ..

 Align seg 1/1  to reverse of: AI218954  from: 1  to: 552

      68 TyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProAr 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     552 TATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCAG 503

      84 gGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspL 101
         ||||||||||      |||   ||||||   ||||||||||||||||||
     502 GCAGAGATTTGGAGACCATGGGCTGGTACAAGATATCAAGGGCGCCGACC 453

     101 euLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPro 117
         |||||:::||||||||||||||||||||||||||||||||||||||||||
     452 TGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCCC 403

     118 SerProProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLe 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCT 353

     134 uGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgG 151
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 TCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTCGAG 303

     151 lySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluPro 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGAACCT 253

     168 GluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLe 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCCT 203

     184 uGlnAlaProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaG 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 CCAGGCTCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCGG 153

     201 luIleAlaAlaCys 205
         ||||||||||||||
     152 AGATAGCTGCCTGC 139

seq_name: gb_est26:AI936064

seq_documentation_block:
LOCUS       AI936064      534 bp    mRNA            EST       08-MAR-2000
DEFINITION  wo47h12.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2458535 3'
            similar to SW:D2_ONCVO P54187 D2 PROTEIN ;, mRNA sequence.
ACCESSION   AI936064
VERSION     AI936064.1  GI:5674934
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 942   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2458535"
                     /clone_lib="NCI_CGAP_Gas4"
                     /tissue_type="poorly differentiated adenocarcinoma with
                     signet ring cell features"
                     /lab_host="DH10B"
                     /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
```

```
                    Average insert size 1.69 kb. Life Technologies catalog #:
                    11549-011"
BASE COUNT       96 a    133 c    165 g    140 t
ORIGIN


alignment_scores:
            Quality:  695.00             Length:    145
              Ratio:  4.964                Gaps:      0
 Percent Similarity:  96.552   Percent Identity: 96.552

alignment_block:
 US-09-215-435-225_COPY_23_227 x AI936064/rev   ..

 Align seg 1/1  to reverse of: AI936064  from: 1  to: 534

      61 GlyAlaValAspGlyAlaThrTyrIleLeuValMetValAspProAspAl 77
         |||||||||   ||||||||||||||||||||||||||||||||||||||
     531 GGGGCCGTGCCGGGCGCAACCTATATCCTGGTGATGGTGGATCCAGATGC 482

      77 aProSerArgAlaGluProArgGlnArgPheTrpArgHisTrpLeuValT 94
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 CCCTAGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAC 432

      94 hrAspIleLysGlyAlaAspLeuLysLysGlyLysIleGlnGlyGlnGlu 110
           |||   ||||||||||||||||||||||||||||||||||||||||||
     431 AAGATAC.AAGGGCGCCGACCTGAAGAAAGGAAAGATTCAGGGCCAGGAG 383

     111 LeuSerAlaTyrGlnAlaProSerProProAlaHisSerGlyPheHisAr 127
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     382 TTATCAGCCTACCAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCG 333

     127 gTyrGlnPhePheValTyrLeuGlnGluGlyLysValIleSerLeuLeuP 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     332 CTACCAGTTCTTTGTCTATCTTCAGGAAGGAAAAGTCATCTCTCTCCTTC 283

     144 roLysGluAsnLysThrArgGlySerTrpLysMetAspArgPheLeuAsn 160
         ||||||||||||||||   ||||||||||||||||||||||||||||||
     282 CCAAGGAAAACAAAACTCCAGGCTCTTGGAAAATGGACAGATTTCTGAAC 233

     161 ArgPheHisLeuGlyGluProGluAlaSerThrGlnPheMetThrGlnAs 177
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     232 CGCTTCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAGAA 183

     177 nTyrGlnAspSerProThrLeuGlnAlaProArgGluArgAlaSerGluP 194
         ||||||||||||||||||||||||||||||||   ||||||||||||||
     182 CTACCAGGACTCACCAACCCTCCAGGCTCCCAGAGGAAGGGCCAGCGAGC 133

     194 roLysHisLysAsnGlnAlaGluIleAlaAlaCys 205
         ||||||||   ||||||||||||||||||||||||
     132 CCAAGCACAA.AACCAGGCAGAGATAGCTGCCTGC 99

seq_name: gb_est93:BF846558

seq_documentation_block:
LOCUS       BF846558      532 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0058-261000-441-a06 EN0058 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846558
VERSION     BF846558.1  GI:12233708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-EN0058-
            261000-441-a06&t3=2000-10-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 29
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0058"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      140 a    158 c    138 g     96 t
ORIGIN


alignment_scores:
            Quality:  687.50             Length:    136
              Ratio:  5.169                Gaps:      1
 Percent Similarity:  97.794   Percent Identity: 95.588

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF846558   ..

 Align seg 1/1  to: BF846558  from: 1  to: 532

      70 LeuValMetValAspProAspAlaProSerArgAlaGluProArgGlnAr 86
         ||||||||||||||   :::   ||||||||||||||||||||||||||
      27 CTGGTGATGGTGGAT...CAGATGCCTAGCAGAGCAGAACCCAGACAGAG 73

      86 gPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAspLeuLysL 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
      74 ATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGACCTGAAGG 123

     103 ysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaProSerPro 119
         ::|||||||||||||||||||||||||||||||||||||||||||||||||
     124 AAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCCCTCCCCA 173

     120 ProAlaHisSerGlyPheHisArgTyrGlnPhePheValTyrLeuGlnGl 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     174 CCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCTTCAGGA 223

     136 uGlyLysValIleSerLeuLeuProLysGluAsnLysThrArgGlySerT 153
         ||||||||||   ||||||||||||||||||||||||||||||||||||||
     224 AGGAAAAGTCACCTCTCTCCTTCCCAAGGAAAACAAAACTCGAGGCTCTT 273

     153 rpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGluProGluAla 169
         ||||||||||||||:::||||||||||||||||||||||||||||||||||
     274 GGAAAATGGACAGACTTCTGAACCGTTTCCACCTGGGCGAACCTGAAGCA 323

     170 SerThrGlnPheMetThrGlnAsnTyrGlnAspSerProThrLeuGlnAl 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     324 AGCACCCAGTTCATGACCCAGAACTACCAGGACTCACCAACCCTCCAGGC 373

     186 aProArgGluArgAlaSerGluProLysHisLysAsnGlnAlaGluIleA 203
```

```
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    374 TCCCAGAGAAAGGGCCAGCGAGCCCAAGCACAAAAACCAGGCGGAGATAG. 423

    203 laAlaCys 205
        ||||||||
    424 CTGCCTGC 431

seq_name: gb_est110:W32197

seq_documentation_block:
LOCUS       W32197         451 bp    mRNA            EST       10-OCT-1996
DEFINITION  zc04f05.r1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            IMAGE:321345 5' similar to SW:PBPL_RAT P31044 23 KD
            MORPHINE-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   W32197
VERSION     W32197.1  GI:1313189
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 705   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Homo sapiens"
                     /db_xref="GDB:1259003"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:321345"
                     /clone_lib="Soares_parathyroid_tumor_NbHPA"
                     /tissue_type="parathyroid tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTT
                     TTTTT-3'], double-stranded cDNA was size selected, ligated
                     to Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                     adenomas was kindly provided by Dr. Stephen Marx, National
                     Institute of Diabetes and Digestive and Kidney Diseases,
                     NIH."
BASE COUNT      124 a    122 c    109 g     90 t      6 others
ORIGIN


alignment_scores:     -
            Quality:  686.00             Length:    152
              Ratio:  4.667                Gaps:      2
 Percent Similarity:  96.711   Percent Identity: 94.079

alignment_block:
```

```
US-09-215-435-225_COPY_23_227 x W32197   ..

Align seg 1/1  to: W32197  from: 1  to: 451

     34 IleGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleTh 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 ATTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCAC 51

     50 rSerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaT 67
        ||||||||||||||||||||||||||||||||||||||||||||   :::|
     52 CTCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACNG.CANA 100

     67 hrTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluPro 83
        ||||||||||||||||||||||||||||||||||||||||||   ||||||
    101 CCTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGANA.GAACCC 149

     84 ArgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAs 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    150 AGACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGA 199

    100 pLeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaP 117
        |||||||:::|||||||||||||||||||||||||||||||||||   |
    200 CCTGAAGGAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGNT.C 248

    117 roSerProProAlaHisSerGlyPheHisArgTyrGln.PhePheValTy 133
        ||||||||||||||||||||||||||||||||||||| ||||||||||||
    249 CCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAAGTTCTTTGTCTA 298

    133 rLeuGlnGluGlyLysValIleSerLeuLeuProLysGluAsnLysThr. 149
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    299 TCTTCAGGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACTC 348

    150 ArgGlySerTrpLysMetAspArgPheLeuAsnArgPheHisLeuGlyGl 166
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    349 NGAGGCTCTTGGAAAATGGACAGATTTCTGAACCGTTTCCACCTGGGCGA 398

    166 uProGluAlaSerThrGlnPheMetThrGlnAsnTyrGlnAspSerProT 183
        ||||||||||||||||||||||||||||||||||||||||||||||   |
    399 ACCTGAAGCAAGCACCCAGTTCATGACCCAGAACTACCAGGACTCANCAA 448

    183 hr 183
        ||
    449 CC 450

seq_name: gb_est93:BF853920

seq_documentation_block:
LOCUS       BF853920       499 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-d01 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853920
VERSION     BF853920.1  GI:12241664
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
```

```
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0093-
            261200-004-d01&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 227.
FEATURES              Location/Qualifiers
     source           1. .499
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="EN0093"
                      /dev_stage="Adult"
                      /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT      119 a    129 c    148 g    103 t
ORIGIN


alignment_scores:
            Quality:  683.00            Length:    127
              Ratio:  5.421               Gaps:      0
 Percent Similarity:  99.213   Percent Identity: 97.638

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF853920   ..

 Align seg 1/1  to: BF853920  from: 1  to: 499

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     114 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 163

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     164 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 213

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     214 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 263

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||   ||||||||||||||||||||||::
     264 TCCTGGATGGAGCCGATAGTCAAGTCCCCGGGGGCCGTGGACGGCGCAAA 313

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         :|||||||||||||||||||||||||||||||||||||||||||||||||||
     314 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 363

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     364 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 413

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         |||||||||||||||||||||||||||||||||:::||||||||||||||
     414 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGGCTACCAGGCTCC 463

     117 oSerProProAlaHisSerGlyPheHisArg 127
         |||||||||||||||||||||||||||||||
     464 CTCCCCACCGGCACACAGTGGCTTCCATCGC 494

seq_name: gb_est93:BF846063

seq_documentation_block:
LOCUS       BF846063      464 bp    mRNA            EST       16-JAN-2001
DEFINITION  QV0-EN0055-181000-439-f11 EN0055 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF846063
VERSION     BF846063.1  GI:12233213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-EN0055-
            181000-439-f11&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 464.
FEATURES              Location/Qualifiers
     source           1. .464
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="EN0055"
                      /dev_stage="Adult"
                      /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       94 a    134 c    122 g    114 t
ORIGIN


alignment_scores:
            Quality:  681.00            Length:    124
              Ratio:  5.492               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF846063/rev   ..

 Align seg 1/1  to reverse of: BF846063  from: 1  to: 464

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     390 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 341

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     340 CACCCTCTTTTGCCAGGGCCTTGAAGTCTTCTACCCAGAGTTGGGGAACA 291

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     290 TTGGCTGCAAGGTTGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 241

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
```

```
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    240 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 191

     67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    190 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 141

     84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    140 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 91

    101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     90 CTGAAGAAAGGGAAGATTCAGGGCCAGGAGTTATCAGCCTACCAGGCTCC 41

    117 oSerProProAlaHisSerGly 124
        ||||||||||||||||||||||
     40 CTCCCCACCGGCACACAGTGGC 19

seq_name: gb_est93:BF853928

seq_documentation_block:
LOCUS       BF853928      536 bp    mRNA            EST       16-JAN-2001
DEFINITION  MR2-EN0093-261200-004-g12 EN0093 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853928
VERSION     BF853928.1  GI:12241672
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-EN0093-
            261200-004-g12&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0093"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      121 a    139 c    164 g    112 t
ORIGIN
```

```
alignment_scores:
           Quality:  676.00              Length:    126
             Ratio:  5.408                 Gaps:      0
 Percent Similarity:  99.206   Percent Identity: 97.619

alignment_block:
 US-09-215-435-225_COPY_23_227 x BF853928   ..

 Align seg 1/1  to: BF853928  from: 1  to: 536

       1 AspGluAspGluAsnSerProCysAlaHisGluAlaLeuLeuAspGluAs 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 GACGAGGATGAGAACAGCCCGTGTGCCCATGAGGCCCTCTTGGACGAGGA 206

      17 pThrLeuPheCysGlnGlyLeuGluValPheTyrProGluLeuGlyAsnI 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 CACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGGGGAACA 256

      34 leGlyCysLysValValProAspCysAsnAsnTyrArgGlnLysIleThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TTGGCTGCAAGGTCGTTCCTGATTGTAACAACTACAGACAGAAGATCACC 306

      51 SerTrpMetGluProIleValLysPheProGlyAlaValAspGlyAlaTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCCTGGATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAAC 356

      67 rTyrIleLeuValMetValAspProAspAlaProSerArgAlaGluProA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 CTATATCCTGGTGATGGTGGATCCAGATGCCCCTAGCAGAGCAGAACCCA 406

      84 rgGlnArgPheTrpArgHisTrpLeuValThrAspIleLysGlyAlaAsp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 GACAGAGATTCTGGAGACATTGGCTGGTAACAGATATCAAGGGCGCCGAC 456

     101 LeuLysLysGlyLysIleGlnGlyGlnGluLeuSerAlaTyrGlnAlaPr 117
         ||||||||||||||||||||||||||||||||||:::  |||:::||||||
     457 CTGAAGAAAGGGAAGATTCAGGGGCAGGAGTTATCAGGCTAACAGGGTCC 506

     117 oSerProProAlaHisSerGlyPheHis 126
         |||||||||||||||||||||||||||
     507 CTCCCCACCGGCACACAGTGGCTTCCAT 534
```